



Calhoun: The NPS Institutional Archive
DSpace Repository

Theses and Dissertations

1. Thesis and Dissertation Collection, all items

2003-06

Using commercial-off-the-shelf speech recognition software for conning U.S. warships

Tamez, Dorothy J.

Monterey, California. Naval Postgraduate School

<http://hdl.handle.net/10945/898>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



<http://www.nps.edu/library>

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THESIS

**USING COMMERCIAL-OFF-THE-SHELF SPEECH
RECOGNITION SOFTWARE FOR CONNING U.S.
WARSHIPS**

by

D. J. Tamez

June 2003

Thesis Advisor:
Co-Advisor:

Monique P. Fargues
Russell Gottfried

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 2003		3. REPORT TYPE AND DATES COVERED Master's Thesis
4. TITLE AND SUBTITLE: Using Commercial-Off-The-Shelf Speech Recognition Software for Conning U.S. Warships			5. FUNDING NUMBERS	
6. AUTHOR(S) D. J. Tamez				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) <p>The U.S. Navy's Transformation Roadmap is leading the fleet in a smaller, faster, and more technologically advanced direction. Smaller platforms and reduced manpower resources create opportunities to fill important positions, including ship-handling control, with technology.</p> <p>This thesis investigates the feasibility of using commercial-off-the-shelf (COTS) speech recognition software (SRS) for conning a Navy ship. Dragon NaturallySpeaking Version 6.0 software and a SHURE wireless microphone were selected for this study. An experiment, with a limited number of subjects, was conducted at the Marine Safety International, San Diego, California ship-handling simulation facility. It measured the software error rate during conning operations. Data analysis sought to determine the types and significant causes of error. Analysis includes factors such as iteration number, subject, scenario, setting and ambient noise. Their significance provides key insights for future experimentation.</p> <p>The selected COTS technology for this study proved promising overcoming irregularities particular to conning, but the software vocabulary and grammar were problematic. The use of SRS for conning ships merits additional research, using a limited lexicon and a modified grammar which supports conning commands. Cooperative research between the Navy and industry could produce the "Helmsman" of the future.</p>				
14. SUBJECT TERMS Speech Recognition, Voice Recognition, Ship Maneuvers, Conning Commands, Seamanship, Ship Handling, Commercial-Off-The-Shelf Software, Dragon NaturallySpeaking Professional Version 6.0, Automatic Speech Recognition			15. NUMBER OF PAGES 92	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

**USING COMMERCIAL-OFF-THE-SHELF SPEECH RECOGNITION
SOFTWARE FOR CONNING U. S. WARSHIPS**

Dorothy J. Tamez
Lieutenant, United States Navy
B.M.Sc., Emory University, 1991

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
June 2003**

Author: Dorothy J. Tamez

Approved by: Monique P. Fargues
Thesis Advisor

Russell Gottfried
Co-Advisor

Dan C. Boger
Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The U.S. Navy's Transformation Roadmap is leading the fleet in a smaller, faster, and more technologically advanced direction. Smaller platforms and reduced manpower resources create opportunities to fill important positions, including ship-handling control, with technology.

This thesis investigates the feasibility of using commercial-off-the-shelf (COTS) speech recognition software (SRS) for conning a Navy ship. Dragon NaturallySpeaking Version 6.0 software and a SHURE wireless microphone were selected for this study. An experiment, with a limited number of subjects, was conducted at the Marine Safety International, San Diego, California ship-handling simulation facility. It measured the software error rate during conning operations. Data analysis sought to determine the types and significant causes of error. Analysis includes factors such as iteration number, subject, scenario, setting and ambient noise. Their significance provides key insights for future experimentation.

The selected COTS technology for this study proved promising overcoming irregularities particular to conning, but the software vocabulary and grammar were problematic. The use of SRS for conning ships merits additional research, using a limited lexicon and a modified grammar which supports conning commands. Cooperative research between the Navy and industry could produce the "Helmsman" of the future.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	VOICE ACTIVATED COMMAND SYSTEM	1
B.	BACKGROUND	1
C.	SIGNIFICANCE TO THE U.S. NAVY	3
D.	MISSION NEED	7
E.	PREVIOUS VACS STUDIES	10
F.	VOICE RECOGNITION TECHNOLOGY	11
II.	SPEECH RECOGNITION SOFTWARE.....	15
A.	SPEECH RECOGNITION SOFTWARE (SRS) COMPONENTS.....	15
B.	INTRODUCTION TO SPEECH RECOGNITION PROCEDURE	16
C.	SPEECH RECOGNITION PARAMETERS	18
D.	DRAGON NATURALLYSPEAKING VERSION 6.0 (DNSV6.0).....	20
E.	SPEECH RECOGNITION HARDWARE REQUIREMENTS	21
F.	VOCABULARY	22
G.	USER ENROLLMENT.....	23
H.	METRICS	24
III.	METHODOLOGY AND DATA COLLECTION FOR VOICE RECOGNITION SOFTWARE EXPERIMENT	27
A.	OVERVIEW	27
B.	EXPERIMENT OBJECTIVES	28
C.	EXPERIMENT DESIGN	28
D.	EQUIPMENT AND SIMULATOR	30
E.	EXPERIMENT SETTING.....	31
F.	EXPERIMENT PROCESS.....	34
G.	EXPECTATION AND CONSIDERATIONS	34
	1. Expectations	34
	2. Considerations.....	36
H.	SUBJECTS	38
IV.	DATA PREPARATION AND ANALYSIS OF EXPERIMENT RESULTS.....	41
A.	EXPERIMENT SCENARIO	41
B.	DATA PREPARATION	43
	1. Data Analysis Requirements	43
	2. Influential Observation	45
	3. Anova Methodology	48
	4. Inference Testing	49
	a. <i>Expectation 1</i>	49
	b. <i>Expectation 2</i>	51
	c. <i>Expectation 3</i>	52
	d. <i>Expectation 4</i>	53
	e. <i>Expectation 5</i>	54

C.	EXPERIMENTAL DESIGN AND IMPLEMENTATION LESSONS LEARNED	57
V.	CONCLUSIONS AND RECOMMENDATIONS.....	59
A.	ANALYTICAL CONCLUSIONS	59
B.	IMPACT OF THIS STUDY	60
C.	RECOMMENDATIONS FOR FUTURE STUDY	61
	APPENDIX A. SHIP-HANDLING VOCABULARY	63
	APPENDIX B. EXPERIMENT RESULTS	65
	APPENDIX C. ACRONYMS	67
	LIST OF REFERENCES.....	71
	INITIAL DISTRIBUTION LIST	75

LIST OF FIGURES

Figure 1.	Simple Voice Activated System.....	2
Figure 2.	Speech Recognition Software Components.....	16
Figure 3.	Hidden Markov Model.....	17
Figure 4.	MSI Console Room.	32
Figure 5.	Quantiles of Standard Normal with Trials 1, 4, and 6 Marked as the Most Significant.	46
Figure 6.	Cook's Distance with Trials 1, 4, and 19 Marked as Significant.....	47
Figure 7.	Trimmed Results Plot.	49
Figure 8.	Subject Error Performance Similarities.....	50
Figure 9.	Trial Comparison.	52

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Common Speech Recognition Parameters.	18
Table 2.	All The Reviewers Reviewed Chart.	20
Table 3.	Experiment Expectations.	28
Table 4.	MSI Noise Levels.	33
Table 5.	Subject Traits.	39
Table 6.	Raw Data Results.	43
Table 7.	Logit Transform Values.	45
Table 8.	P-Value for Expectation 1.	50
Table 9.	P-Value for Expectation 2.	52
Table 10.	P-Value for Expectation 3.	53
Table 11.	Ninety-Five Percent Confidence Interval ($t = 2.086$).	53
Table 12.	P-Value for Expectation 5 (Full Model Including Scenario-Subject Interaction).	55
Table 13.	P-Value for Expectation 5 (Subject * Setting).	56
Table 14.	P-Value Expectation 5 (Setting * Trial).	56
Table 15.	Error Types.	65
Table 16.	Conditions Per Subject Per Trial.	65

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I thank God for this scholastic opportunity and for giving me the knowledge, strength, and academic support to complete this thesis. Only through His will did this thesis come to fruition.

I thank my thesis advisors, Dr. Monique Fargues and LCDR Russell Gottfried, for their time, energy, patience, and most of all their expertise. Their insight and guidance are most appreciated. LCDR Gottfried's statistical expertise provided analytical granularity, focusing the findings and clarifying the results of the experiment. I am grateful to have been blessed with such outstanding advisors.

I send warm wishes and much thanks to Admiral Ramsey and his staff at Marine Safety International, San Diego, California. The staff's encouragement, participation, and support enhanced the experiment. I especially thank Don, Rob, Garry and Bill for their participation in the experiment and their Navy ship-handling expertise and Doug for his computer savvy and technical abilities.

I thank my family and friends, in particular Ginifer and Rich for always being there when needed. Last, I dedicate this thesis to the loving memory of my parents, Raúl and Gail, for their loving support and encouragement.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. VOICE ACTIVATED COMMAND SYSTEM

This thesis focuses on speech exchanges during ship control processes and specifically considers the potential of Commercial-Off-The-Shelf (COTS) voice recognition software as part of a Voice Activated Command System (VACS) to replace Sailors in this process. VACS is a complex, multifaceted, automated system designed to perform the functions of a Helmsman who adjusts the ship's rudder angle, and a Lee Helmsman who adjusts the ship's engine speed. The VACS uses speech recognition software to identify and transmit the Conning Officer's commands to software programs interfacing with the rudder and engines.

Voice recognition, also referred to as speech recognition (SR), software is a vital part of the VACS. The rudder and engine applications would rely on accurate input from the voice recognition software. Commercial-Off-The-Shelf (COTS) voice recognition software is currently available for evaluation and a prospective technology for conning U.S. warships. This study reviews the potential strengths and weaknesses, design considerations and recommendations for future research of the selected software in a Voice Activated Command System.

B. BACKGROUND

Speech has been for centuries and is today the primary form of communication in controlling ship's maneuvers. Speech can be used at a distance which makes it ideal for hands-busy and eyes-busy situations. The enduring truth about verbal communication is that the receiver, a Helmsman, must successfully interpret the information passed from the person responsible for maneuvering the ship. The message or command must be clear and concise using a vocabulary common to both parties.

During the 17th and 18th centuries, the ship's Captain ordered adjustment of the sails to gain speed. He passed a verbal order down the chain of command and the appropriate Sailor changed the rigging. Later the Captain delegated these duties to Conning Officers, responsible for ordering shipboard maneuvering. Regardless of technological improvement in exchanging important information via wireless computers using Voice over Internet Protocol, ship maneuvering dynamics have not changed. A Conning Officer still voices commands to a Helmsman who converts it to action. Changes in transmission media have led to more effective, convenient or efficient processes of performing key tasks. These changes include the development of Voice Activated Systems (VAS), Figure 1, computer software that activates machines using the human voice. Speech recognition software transforms sound waves from voice into digital bits. An interface then interprets them as commands and converts them to mechanical or electrical signals. Resulting signals are relayed to the rudder and engine to adjust the angle and speed accordingly.

VAS Diagram

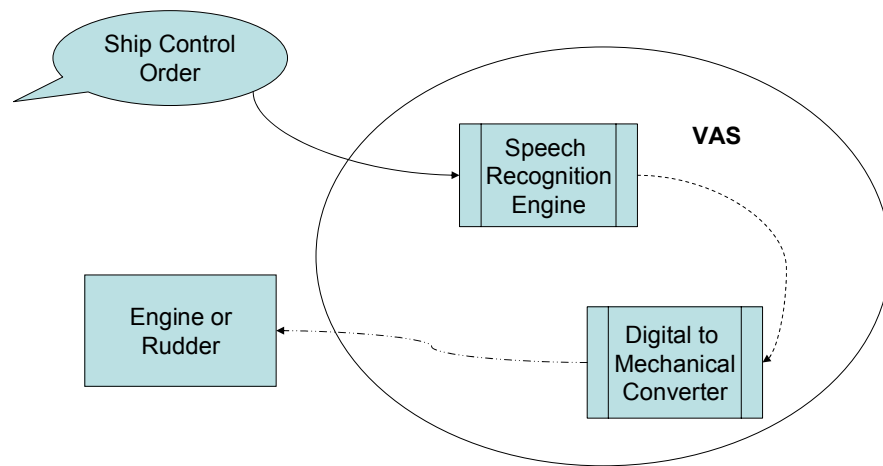


Figure 1. Simple Voice Activated System.

Over the last decade the use of VAS has become more common and in greater demand. Voice Activated Systems are most common in the telephone industry, but as the technology matures their use spreads to new areas. The technology routinely responds to people speaking key-words, telephones dial a caller's spoken number or allows businesses to automate transactions via computer generated dialogues. Persons with disabilities are gaining personal freedom and a sense of accomplishment by using Voice-activated Environmental Control Units, which enable them to control a full range of electrical household items simply by giving verbal commands. [Ref. 1] The same voice technology that initiates turning on and off lights or alarm systems can make a valuable contribution to Navy systems.

Driving or conning a ship is a prime example of human interaction, which evolved around and through speech and where a Voice Activated System could be instrumental. The Conning Officer gives a standardized verbal command and the Helmsman or Lee Helmsman responds with a formal verbal acknowledgement and then a verbal update of the ship's status. To conclude the sequence the Conning Officer states an understanding of the ship's status. Conning a ship is manpower intensive and subject to human error, which VAS may assist in alleviating.

C. SIGNIFICANCE TO THE U.S. NAVY

The U.S. Navy faces numerous challenges now and in the future and stands at the threshold of numerous significant changes. "Our goal is to move our military from service-centric forces armed with unguided munitions and combat formations that are large and easily observable, manpower intensive, earth-bound capabilities, and transform a growing portion into rapidly-deployable joint forces made up of less manpower intensive combat formations...." [Ref. 2]

One of the most apparent and serious challenges is how to perform all the mission requirements with a smaller force. Manpower reductions occurred steadily throughout the 1990's creating personnel shortages on naval platforms.

To meet its future objectives, the Navy is evaluating methods to reduce manning on each platform so that more ships may be put into service without increasing overall personnel end strength. An increased number of smaller, less manpower intensive ships may be dispersed across multiple theaters simultaneously. These ships would fill different mission requirements to meet the multitude of diverse threats to U.S. interests. Innovative techniques to reduce ship manning, without sacrificing readiness or jeopardizing the mission greatly, benefit the Navy, especially since manpower-related expenses combine to consume approximately 60% of its budget. [Ref. 3]

Department of Defense (DoD) and Navy leaders seek less expensive, more productive and effective approaches to resolve this issue. The Secretary of the Navy stated that one immediate goal is to “explore innovative manning initiatives such as the Optimum Manning program, which relies on new technologies and creative leadership to reduce ship manning.” [Ref. 4] Optimal Manning program prototypes are in place aboard the USS MILIUS and the USS MOBILE BAY. On board MILIUS, the Optimum Manning program, part of the Smart Ship concept, is operating with an “optimal crew size of just 232, almost 20% less crew than the usual complement for an Arleigh Burke-class guided missile destroyer.” [Ref. 5] MILIUS and MOBILE BAY report success using an optimal crew by introducing new technology and new policies and procedures, characteristic of the Navy’s transformation. The advances on these ships open the doors for Navy officials to research the feasibility of designing new ships and retro-fitting current ships with VACS.

The Voice Activated Command System (VACS) has the capability to reduce shipboard watch standing and maintenance manpower requirements. VACS may substitute for the Helm and Lee Helm positions. On smaller platforms this means the elimination of at least a single watch-stander, but as many as three watch-standers: the Helm, Lee Helm and the Helm Safety Officer. This reduction enables redistribution of less skilled roles to highly skilled technical or decision making billets on board a warship, such as the Littoral Combat Ship (LCS).

The Navy's future LCS is a multi-mission surface combatant designed for operation within 100 miles of land. LCS concepts require a smaller, faster and more versatile vessel than its predecessors. The manning in LCS is projected to be severely reduced compared to current day standards. The smaller crew emphasizes the need to ensure every possible member is performing mission critical tasks. One assumption regarding the design of the LCS is that it will leverage as much technology as possible to meet the proposed manning level.

Currently, the Helm watch is posted twenty-four hours a day, seven days a week while a ship is underway. This manning would necessitate three helmsmen on eight hour shifts without any time off. Given a 35-50 man crew with a helmsman working an eight hour shift and the helm manned 24 hours a day, seven days a week, approximately six to eight and a half percent of the crew drives the ship full time, not including the conning officer. Manning at reduced levels risks fatigue, provides little redundancy and leaves no room for training personnel for replacement. Increased manning for this watch station would require more helmsmen, as much as doubling the manpower requirements.

Navy leaders and ship designers are presently exploring technological alternatives to reduce shipboard manning requirements. One potential area includes VACS to interact with the Ship System Control segment of the Integrated Bridge on the Littoral Combat Ship to help reduce manning. For example, use of VACS aboard LCS would eliminate the Helmsman watch station allowing a significant portion of the crew to concentrate on performing other more skilled duties. The deployment of a well designed, technologically advanced LCS will greatly enhance Littoral Sea Control and assist in the Navy's transformational programs.

The Naval Transformation Roadmap (NTR) and Joint Vision 2020 (JV 2020) describe strategies, concepts, initiatives and programs considered crucial in transforming the Department of Defense and the Navy in particular. The following quote emphasizes the need for technologically advanced, automated warships such as the LCS.

This transformation is motivated by a vastly different security environment that has emerged over the last decade. Where once a single monolithic threat—the Soviet Union—dominated the nation’s security planning and programming, today’s environment contains a broader, more diffuse set of concerns: terrorism, biological warfare, regional tension, and an array of other transnational challenges. [Ref. 6]

As stated previously, the need for an LCS drives the need for VACS. Both NTR and JV 2020 stress the Navy’s need for interagency cooperation and technological change. One major theme communicated in the Transformation Roadmap is “...inserting technology to carry out operations in ways that profoundly improve current capabilities and develop desired future capabilities.” [Ref. 7] VACS fulfills that requirement. It can offer an effective and less manpower intensive option for maneuvering a ship, to which personnel can relate and adapt quickly, with minimal disruption to the current modis operandi.

Essentially, the technical and operational transition can be made because the VAC system may be designed to use the same inputs as a human helmsman. Experimentation must demonstrate that VACS software ensures conning commands are delivered in the correct format and that the order given is the most appropriate for the intended maneuver. Unlike people, a computer does not interpret commands delivered in the incorrect format, nor does it make adjustments for orders that do not do exactly match what the Conning Officer intended. Conning officers need to use the standard command set to match the system’s predefined vocabulary.

The system assists with future capabilities as part of the FORCEnet architecture, an all-inclusive maritime network intended to provide combatants all necessary information and support in real-time. As an integral part of the Littoral Combat Ship, VACS supports the Sea Shield and Homeland Defense strategies. The utilization of smaller, more agile craft with smaller crew size and the need for reliability and precision make VACS a strong candidate solution for fleet

operations. VACS can reduce manpower need, therefore reducing the number of Service personnel exposed to “decisive points in battle or in other operations, or to be exposed to conditions of great danger and hardship”. [Ref 8]

D. MISSION NEED

State of the art Ship Control System that makes efficient use of technology enables improved command and control of U.S. Navy surface vessels and diverts manning to other shipboard war-fighting requirements. The main objective of a Voice Activated Command System is to replace the helmsman and lee helmsman. VACS is aimed at responding to conning commands in the same manner as a helmsman, providing feedback, updates and performing its primary mission of transmitting the appropriate control signals to the rudder or engine.

The Voice Activated Command System must meet four overarching criteria: reliability, multiple-user capability, speaker verification and noise dampening capability. Each of these criteria is vital for use on a U.S. warship to ensure additional complications do not occur due to malfunctioning software, misinterpretation of commands or simply missing orders to the helm, especially considering the inherent dangers and hazards associated with shipboard maneuvering.

Reliability is defined as the capacity of the VACS to recognize and accurately relay commands. The level of confidence for reliability and accuracy for this system must be near perfect. Ship handler confidence in system operability is essential. Full confidence in the software leads to operational implementation. Use of unproven technology invites unnecessary risks. Technology determined to be unreliable collects dust while Sailors continue to use antiquated, more costly, but proven processes. Most important, even momentary system failure could result in harm to the ship or crew, costing millions of dollars in repairs, or worse, Sailors’ lives.

Ship control duties rotate among multiple users and must quickly and smoothly transition from one user to another. The VRS software must recognize the speech patterns, inflections and accents of each individual user. Several different conning officers assume the Watch on each ship, creating the need for accommodating a pool of watch-standers, one at a time. The watches are set for limited periods of time to ensure awareness and to reduce mental and physical fatigue. These factors increase the number of VACS users, thereby increasing the need for the software to accurately respond to multiple users. The ability to respond to a number of distinct users must be balanced by the requirement to accept only the responsible individual's command.

Speaker verification or authentication guarantees the VACS software only listens to the authorized Conning Officer on watch. In addition to the Conning Officer, an Officer of the Deck (OOD) oversees all maneuvering and seamanship duties. The OOD is the Commanding Officer's direct representative and the VACS must be programmed to respond to an emergency order from the OOD or to disregard that voice, even if stating a standard command, and only acknowledge and execute the commands from the currently authorized Conning Officer. Speaker verification also allows for user permissions to be set, such as a hierarchy of emergency or safety overrides. The Commanding Officer and Executive Officer require the ability to negate, interrupt or override commands given by officers with subordinate permissions. As specified by regulation or standards, officers with more qualified permissions may be allowed to interrupt or override commands given by subordinate officers as well. Based on the current hierarchical structure, most officers would not be allowed to override any other officer.

The Voice Activated Command System has a few constraints associated with its implementation. The system requires each user to record voice and speech patterns prior to use, thereby training it to understand specific voices stored in its database. The system will respond solely to their voices. The logistics of installing and maintaining the system will require information system technicians are available at all times, in case of emergency. Another crucial

element is the need for a manual over-ride system, a back-up system and alternate power supply in response to malfunctions or emergencies which could prevent proper operation. The last constraint is the operational environment. Like any other system it requires sufficient casing to ensure that the weather (i.e., salt air, lightning strikes or other such problems) does not affect the circuitry. Finally, more than other modalities, there is the possibility of anthropomorphism when using speech recognition. It has been documented that users tend to overestimate the capabilities of a system if a speech interface is used and that users are more tempted to treat the device as another person. [Ref. 9]

Alternatives to VACS are interesting but have significant drawbacks. One option is to not install VACS and maintain the status quo, but this does not allow for reduction of manpower established in the Navy's plans and vision. A Non-voice Activated Command System (NACS) requires the operator to input the data manually. There are three designs under consideration, a console, a wrist watch or a helmet. The primary drawback to the NACS system is that it does not mirror the current process. Conning Officers would have to learn a new process to use any form of this system. Also, other watchstanders or supervisors, including the OOD would not be able to see or hear the command until it is initiated, making it impossible for them to intervene in a timely manner. Console option requires the Conning Officer to remain in a stationary position, which prevents checking the bridge wings or moving about to consult other watch-standers. The wristwatch option is more portable, but requires great dexterity to input the data via a key pad on the wrist, which becomes even more difficult during rough seas, or during close maneuvering operations requiring their full attention. The helmet option would turn the ship based on the wearer's movements. It was initially designed for aviators who remain seated throughout their mission. The helmet is impractical for a conning officer whose safety duties demand motion whenever needed. SRS is the only option that enables immediate oversight and, if necessary, override by senior personnel.

E. PREVIOUS VACS STUDIES

Automating many ship's functions has long been sought by Navy leaders. Periodic experiments have been initiated to determine if technology has developed enough to satisfy the ideas and theories of automating the ship-handling. Conning system automation shows a great deal of promise. Speech recognition technology, considered to be the single greatest hindrance, has significantly improved over the last decade and the Navy's manpower reduction initiatives have necessitated alternatives for executing tasks previously performed by Sailors.

A Voice Activated Command System was tested as part of the Integrated Bridge System HIS Test (DT-IB 509) experiment. [Ref. 10] Preliminary experiment results include the following:

- Enhance Conning Officer situational awareness and ship safety,
- Require high degree of user confidence in accuracy to reduce watch-stander stressors,
- Replicate current verbal ship-handling commands,
- Need standard command vocabulary,
- Need no greater than 0.1 second delay between the command receipt and execution,
- Need less cumbersome support equipment,
- Increase Conning Officer's receptiveness to participating in the experiment,
- Need capability for Conning Officer to take direct control,
- Need displays showing actual position,
- Need ability to vary confidence level for each user,
- Need misinterpretation fixed so that VACS does not take the wrong action or no action at all,
- Participants preferred VACS to NACS.

These initial results demonstrate the promise of technology and principal areas of interest from the Navy in directing research efforts in future experiments. This thesis will focus on speech recognition software accuracy, including

experimentation implementing the use of standard commands. The experiment will not focus on the VACS as a whole. Therefore, the signal transition from digital to mechanical will be tested.

F. VOICE RECOGNITION TECHNOLOGY

The Voice Recognition Industry is growing rapidly as speech is incorporated into more and more applications. The first Automatic Speech Recognition (ASR) system was developed in 1952 at the Bell Laboratories, when it could recognize the numbers zero through nine. Since then, ASR systems have made significant strides and have vocabularies that recognize thousands of words.

There are three main application areas for speech: control and data input in a “hands busy” environment, feedback in visually limited environments, and system control via telephone lines. [Ref. 11] Initially, speech was used mainly for company call centers. Today, speech is becoming commonplace in the home, car and at work, enabling users to interact with people, to control consumer appliances and to access personal and public information. There are toys that interact with children, promoting essential cognitive and motor skills. In automobiles, drivers may request directions and the system tells drivers exact directions from one location to another. With this technology, drivers can change the settings for numerous subsystems using voice commands in some cars.

Voice Activated Command Systems are becoming a greater part of everyday life. One industry group estimates licensing revenues and associated technical proliferation to increase 30-fold between 2002 and 2006. [Ref. 12] One interesting VACS, called e-medICS, allows paramedics to dictate nursing notes and receive life-saving information from the medical facility while on scene. “Being able to operate the e-medICS system by speech commands leaves paramedics' hands free to effect treatment and operates equipment, thus saving vital minutes in the delivery of pre-hospital care”, according to a speech recognition case study. [Ref. 13]

In a draft Request For Proposal (RFP), the U.S. Navy requested new Navigation, Seamanship and Ship-handling Trainers be scalable, to include speech recognition as early as Fiscal Year 2004. The Navy proposes “The voice recognition technology would have the computer respond to all student commands with the appropriate voice response and ship control response” [Ref. 14] in the simulated environment. This request clearly indicates the Navy’s interest in, and intention to, incorporate speech recognition technology into the bridge environment.

Not only is the technology developing but so are the standards which regulate the voice recognition technology. The National Institute of Technology, Speech Group [Ref. 15] is working with the World Wide Web Consortium (W3C) [Ref. 16] to develop baseline standards for voice solutions. These standards lay the foundation for future development. Vendors add proprietary extensions to their products, but the components are built on the same technology, enabling greater interoperability across components and businesses.

Voice Extended Mark-up Language (XML) and Speech Application Language Tags (SALT), voice interface frameworks, are in the final stages of the voice browser certification process. VXML and SALT allow easier implementation of voice applications. Each component is independently evaluated on several technical aspects. Standards are released periodically to help developers plan the progress of a product. This is significant in that standards make the technology more financially and scientifically competitive, create a greater body of knowledge, increase use of the technology and promote collaboration between companies. As product standardization spreads, usually the use increases and the cost decrease. The process enables certification of technicians and engineers for troubleshooting and repairing products, increasing the support base. Another reasonable expectation is that products withstanding the rigors of standards testing would have a longer shelf life. Industry initiatives point in a beneficial direction for developers and consumers and lead the way in establishing a firm technological base for military application of this technology.

Speech Recognition Software has the potential to change how U.S. Navy warships are driven in the future, which will be examined in the following chapters. Chapter II discusses the main concepts behind the speech recognition components. It also presents a brief overview of the speech recognition technology, and specifically Dragon NaturallySpeaking Version 6.0, and defines the metrics used in analyzing this system. Chapter III discusses the experiment equipment, setting, subjects and process considered in this work. The results of the experiment are presented in Chapter IV along with lessons learned about the experimental process. Chapter V covers the conclusions about the experiment and submits recommendations for further research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. SPEECH RECOGNITION SOFTWARE

A. SPEECH RECOGNITION SOFTWARE (SRS) COMPONENTS

Speech recognition is the process of converting an acoustic signal, captured by a microphone into a set of words, and applications can be found, for instance in command and control, data entry, and document preparation. Recognition is usually more difficult when vocabularies are large or have many similar-sounding words. For example, true homonyms within the vocabulary may cause great difficulty for the recognizer. [Ref. 17] The words 'for' and 'four' sound identical yet have very different meanings. The basic recognizer cannot tell which word the user intended. Therefore, several additional specialized components are necessary to recognize human speech, which include the grammar, lexicon, and probabilities based on the user's profile.

Grammars or language models are used to restrict the possible combination of words when speech is produced in a sequence of words. In the 'for' versus 'four' example, the grammar checks the context to determine which word to insert. The lexicon defines the various pronunciations of a word. All components are essential in creating the most accurate speech recognition software, as poor performance by any component severely degrades the overall recognition accuracy rate.

Figure 2 presents the typical components included in a SRS. First, the digitized speech signal is transformed into a set of useful measurements or representations at a fixed rate, typically once every 10 to 20 msec. [Ref. 18] Representations attempt to compactly preserve the information needed to determine the phonetic identity of a sequence of speech while being as impervious as possible to factors such as speaker differences, effects introduced by communications channels, and paralinguistic factors such as the emotional state of the speaker. Representations used in current speech recognizers

concentrate primarily on properties of the speech signal attributable to the shape of the vocal tract rather than to the excitation, whether generated by a vocal-tract constriction or by the larynx, increasing the accuracy.

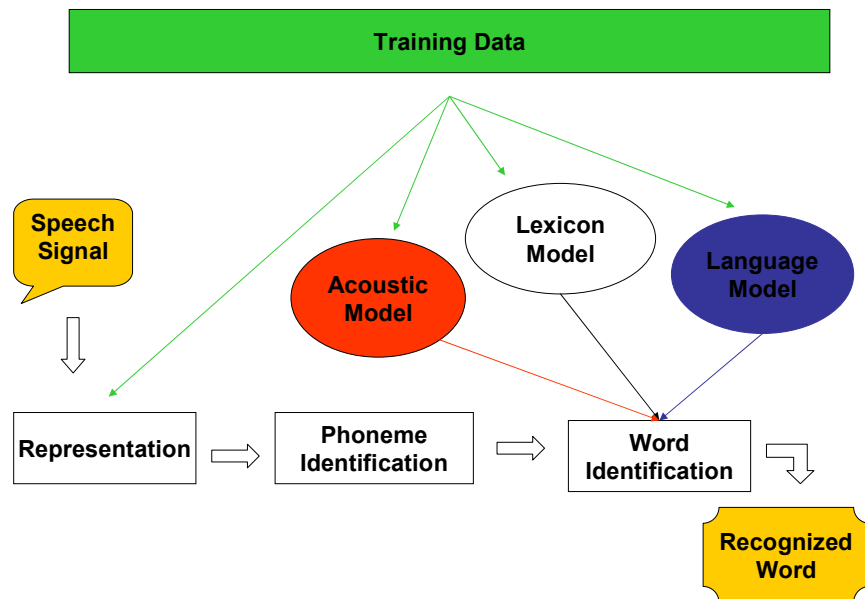


Figure 2. Speech Recognition Software Components.

Next, the resultant measurements are used to search for the most likely word candidate, making use of constraints imposed by the acoustic, lexical, and language models and the training data. Statistical language models, based on estimated frequency of word sequence occurrences are often used to guide the search through the most probable sequence of words.

B. INTRODUCTION TO SPEECH RECOGNITION PROCEDURE

The process of transforming acoustic sounds into written words or commands is complex. The previous section described each component. This section briefly describes how the Automatic Speech Recognition (ASR), grammar and lexicon make the transformation.

The dominant recognition paradigm used for ASR is based on the hidden Markov models (HMM), as illustrated in Figure 3. A hidden Markov model uses a doubly stochastic model, meaning that both the phoneme string (the grammar) and the acoustics (acoustic model) are represented probabilistically as Markov processes. [Ref. 19] The acoustic model captures the acoustic speech properties and provides the probability of the observed acoustic signal given a hypothesized word sequence which includes acoustic analysis and an acoustic model. The acoustic analysis divides the speech into a sequence of acoustic vectors. The acoustic model consists of sub-words called phonemes, which are context dependent and the pronunciation lexicon, which defines the decomposition of the words into the subword units. [Ref. 20].

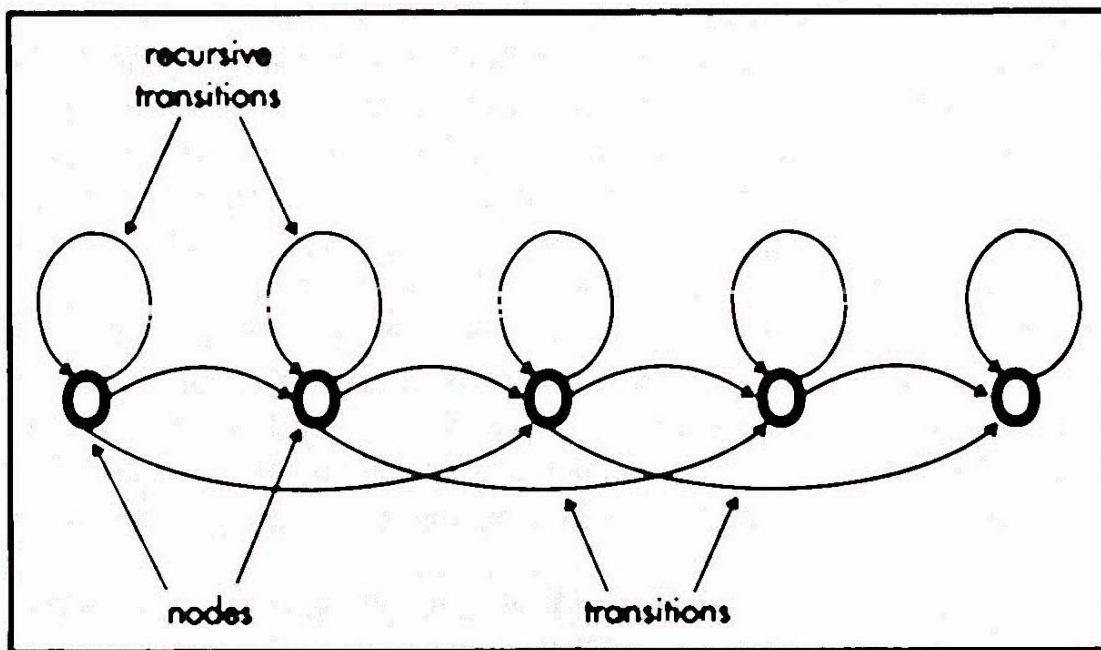


Figure 3. Hidden Markov Model copied with J. Markowitz's consent [Ref. 21].

The grammar or language model provides a statistical estimate for the prior probability of the string of words. N-gram analysis calculates the probability of a given series of words. That is, given the first word of a pair, how confidently can the next word be predicted? [Ref. 22] An N-gram can be viewed as a moving window over a text, where N is the number of words in the window. For

example, Bigrams have two consecutive words, Trigrams: three consecutive words, Quadrigrams: four consecutive words, etc. Words or phonemes have different sounds based on their position in a sentence, emphasizing the need for quality grammars and lexicons.

A lexicon defines the pronunciation of a word and includes information such as phoneme length. It usually includes multiple pronunciations of a word in order to accommodate a wider variety of speech patterns. For example: tomato can be pronounced 'to may to' or 'to mah to'. Lexical design entails two main phases: first, selection of the vocabulary and second, representation of the pronunciation entry using the basic units of the recognition system. Lexicons are often manually created and make use of knowledge and expertise that is difficult to codify. [Ref. 23]

C. SPEECH RECOGNITION PARAMETERS

A criterion used to determine the usefulness or applicability of a SRS to a particular process is called a parameter. Each parameter has a range or scale by which it is measured. The range describes the least to the most complex mode of a specific parameter. Many parameters must be considered when choosing a SRS. Table 1 presents the most common parameters. User adoption rates, environment, amount of training necessary and the accuracy rate are all influenced by the parameters.

PARAMETER	RANGE
Speaking Mode	Isolated Word to Continuous Speech
Speaking Style	Script to Spontaneous
Enrollment	Speaker Dependent to Speaker Independent
Vocabulary	Small (<20 words) to Large (>20,000 words)
Language Model	Finite State to Context Sensitive

Table 1. Common Speech Recognition Parameters.

An isolated-word speech recognition system requires the speaker to pause briefly between words, whereas a continuous speech recognition system allows people to speak more naturally. Spontaneous speech contains speech irregularities, such as 'uhs' and 'ums' and is much more difficult to recognize than speech read from a script. Some software requires speaker enrollment, where a user trains the software by providing speech samples, called a user profile. This training phase allows the system to more easily detect words from background noise, thereby decreasing the error rate. Other SRS are categorized as speaker-independent, in that no enrollment is necessary. Speaker independent software leads to a higher number of errors. In addition, the size of the vocabulary impacts the time necessary to recognize a word. The larger the vocabulary, the longer it may take to recognize it. Finally, a context sensitive language model is more accurate than a finite model. The context sensitive model examines the surrounding words as well as the phonemes to determine the most appropriate word, whereas the finite model makes its determination based solely on the phonemes themselves.

Speech Recognition Software is typically designed for use with a particular set of words, but SRS users may want or need to use words not built into the default vocabulary, leading to out-of-vocabulary word problems. A word not listed in the vocabulary is mapped to a word in the dictionary, causing an error. ScanSoft designed Dragon NaturallySpeaking Version 6.0 to address that problem and other issues arising when using COTS SRS for conning a ship. A SRS must meet certain criteria for use on a U.S. war ship:

- Accuracy rate equal to or greater than a human,
- Ability to respond using verbal ship-handling vocabulary,
- Use standard conning commands,
- Maneuverable support equipment, and
- Concise seamanship vocabulary.

D. DRAGON NATURALLYSPEAKING VERSION 6.0 (DNSV6.0)

DNSV6.0 Professional is a commercial-off-the-shelf (COTS) continuous speech recognition software program designed for use in an office environment. NaturallySpeaking 6 is fast and responsive; it reacts crisply and quickly to both voice commands and dictation. [Ref. 24] The *Consumer Reviewer* “Consensus Report, Table 2, shows the number of times products are top-ranked by reviewers included in All The Reviews Reviewed chart” [Ref. 25] presenting a convincing argument that software and computer reviewers believe DNSV6.0 to be the preferred SRS on the 2002 market. The following characteristics made it appropriate to use for the current study:

# of Picks	Software Brand
7	ScanSoft Dragon NaturallySpeaking Preferred
1	IBM ViaVoice Pro
1	L&H Voice Express (discontinued)

Table 2. All The Reviewers Reviewed Chart.

- A large vocabulary,
- Speaker dependent, indicating greater accuracy,
- Training is quick and easy. A very good speech profile can be created within 15 minutes. An additional 15 to 30 minutes of training leads to an excellent speech profile.
- A centralized accuracy center allowing the user to input their specific information for greater recognition. It has the capability to learn grammatical style and new vocabulary from previously type written documents.
- Ability to handle spontaneous speech and to add words to the vocabulary. The ability to add words is crucial since seamanship terms are not part of the average office environment conversation.
- Capacity to correct the document as the person is speaking

- Highest recognition rate listed among the SRS competitors add to its appeal.
- Ease of use with various computer configurations also made it a logical choice. Z. M. Gao claims one competitor is practically unusable in programs other than Microsoft Word and SpeakPad. [Ref. 26]
- Designed to give commands indicating developers were already researching speech activated command and control.
- Its manufacturer has developed specialized versions for the legal, medical and public works communities, signifying a more easily specialized version for seamanship terms. Some systems are strictly telephony-based and are not well suited to our application.

E. SPEECH RECOGNITION HARDWARE REQUIREMENTS

DNSV6.0 requires the following hardware and software to operate properly in an office setting: Intel® Pentium® II 400 MHz processor, 128 MB RAM, 300 MB free hard disk space, Microsoft® Windows® XP, Millennium, 2000, or 98, a 16- bit recording sound card, Microsoft® Internet Explorer® 5 or higher, a CD-ROM drive, a noise canceling headset microphone and speakers. The speakers allow the other officers on the bridge to hear the system text-to-speech (TTS) responses and confirm the ship's movements. Install DNSV6.0 as a stand-alone application or turn off all software applications not needed, including background applications such as anti-virus detectors. This allows DNSV6.0 to utilize all available computing power and improves recognition accuracy. Although DNSV6.0 works with all these systems, optimal performance is achieved with a 500 MHz processor or faster and 256 MB RAM. [Ref. 27]

There are other criteria that help with the performance when choosing the hardware for this system. Note that the sound card should be of high quality and should have a sound booster, as the sound booster will adjust the sound volume. One tactic frequently used is to turn on the system and not speak for a few seconds. The lack of sound will automatically activate the sound booster, improving recognition accuracy. In addition, close attention should be given to microphone selection. Several sound cards, microphones and speakers are

listed on the manufacturer's web site, which are compatible with the system. Specific consideration regarding the environment, the noise canceling or dampening capability, the user's comfort and the portability (wired versus wireless) of the microphone went into the selection process for the experiment. The Conning Officer's need to move to various stations in and around the bridge greatly restricted the selection to wireless microphones.

Noise dampening capability makes a vast difference in the overall performance of VACS by reducing noise interference from various sources. Noise comes from several sources including the ship's engine or mechanical gear, environmental factors such as wind and rain, co-workers and other bridge equipment. Ships are also known to shudder at times, also contributing to ambient noise. Most of the sources are uncontrollable; therefore the noise dampening capability of VACS becomes more imperative. As a result, the more clearly the acoustics are delivered to the speech recognition software, the greater the resulting accuracy is.

F. VOCABULARY

The global vocabulary in the DNSV6.0 is designed for use by office professionals, who each have their own copy. It is deemed to be large with over 200,000 words. A large vocabulary allows more spontaneous speech with fewer corrections, if the user is stating verbiage typically used in an office. Software designers envisioned one person installing DNSV6.0 at their personnel workstation and then tailoring it for their particular needs, where the tailoring occurs as each user adds words to his/her personal profile. Although adding words seems simple, in reality, it is time consuming because each user must update a personal profile vice one administrator updating the global vocabulary. Also, words cannot be deleted from the global vocabulary. Words that are irrelevant or similar to terms more commonly used by the Conning Officer are compared to the incoming acoustic stream, slowing down the response time and causing errors. Advanced users may overcome this problem by selecting an

Empty Dictation at initial set-up and populating the vocabulary from scratch. [Ref. 28] However, the software was used in the preset configuration since this experiment is designed for novice subjects.

The seamanship vocabulary and the use of DNSV6.0 on board a ship is a challenge for any COTS SRS. Neither DNSV6.0 nor any other current commercial SRS includes seamanship terms in the global vocabulary. The vocabulary is statistically weighted to recall more frequently used words first resulting in new words having a lower statistical rating than words initially listed in the global vocabulary.

The lack of written conning command documentation available to scan into DNSV6.0 to assist learning new words and phrases means the software must learn from current user interaction. DNSV6.0 ability to add words to a user's profile helps immensely in overcoming this problem, as only with repeated use can the SRS learn and recall the seamanship terms prior to words more commonly used in the office environment.

The language used by Conning Officer is unique but standardized. The vocabulary is restricted with approximately one-hundred different words used to drive the ship. The words are set into a strict grammar used for specific maneuvers, called commands.

Even though the phrases are short and standardized there are several ways to pronounce them and minute changes to the phraseology depending on the ship or even on the Commanding Officer. For example, the conning officer may say 'rudder' or 'rudders' amidships on ships with more than one rudder. The 's' on the end seems trivial to the helmsman but the software is not expecting that 's' and looks for a similar word ending in 's', creating an error.

G. USER ENROLLMENT

One reason for choosing DNSV6.0 was due to its easy enrollment as mentioned previously. The system provides step-by-step instructions for every

new user to assist in creating a profile and performing basic functions. The average novice can enroll in approximately 15 minutes. During the enrollment process, the system adjusts the volume setting based on the individual's speaking style. It also evaluates the sound system providing a Speech-to-Noise ratio. Finally, the system records the user's speech pattern and style as he/she reads a set passage.

Speech impediments and an extremely noisy setting will affect the software's ability to complete the user profile and decrease its accuracy rate. Lisps, slurring words, and such will decrease the software's ability to recognize the user's speech. If there are any changes to a person's speaking ability they will need to re-enroll in the system or avoid using it until their voice returns to normal. The optimal setting is a quiet room without any distractions. But, in reality the setting should be similar to the environment in which the software will be used, as background noise in the primary setting will cause distortions if not accounted for during training.

H. METRICS

Error rate or accuracy rate is a common measure used to evaluate SRS performance. Error rate, E is typically described in terms of word error rate and is described in Equation (1) as:

$$E=(S+I+D/N)*100, \quad (1)$$

where, N is the total number of words in the test set, and S , I , and D are the total number of substitutions, insertions, and deletions, respectively. [Ref. 29]

This system's effectiveness has several metrics. Equation (1) will be used to determine the software and the human's accuracy. There are four types of software errors.

- **Software Recognizes the Wrong Word When the Correct Word Is in the Vocabulary**

This is an example of outright misinterpretation. The user may have stated the word differently when creating a user profile. There is a variety of reasons including new context or position in a sentence or different intonation or emphasis on a syllable.

- **Software recognizes the wrong word when the correct word is NOT in the vocabulary**

This is an example of a user stating a word that the software does not have in its vocabulary. The software maps to the word most closely resembling one that is in the vocabulary.

- **Software does not acknowledge a word spoken by the Conning Officer**

This is an example of the software not hearing the word, or hearing it and determining it to be part of another word or background noise. For example, the phrase 'meet her' may be misinterpreted as 'meter'.

- **Software adds a word NOT spoken by the Conning Officer**

This is an example of the language model trying to make the acoustic input into a complete sentence. For example the conning commands state, "steer course 015". The software tries to interpret the sound and follow the grammar built into the software by adding the word 'to' so that the phrase read "steer course to 015".

Along with the software errors there are also human errors in the conning process. There are numerous reasons why a Helmsman may make such an error: distraction, could not hear well or by rote. The helmsman is so accustomed to a particular maneuver in a specific situation and reacts without fully comprehending the Conning Officer's command.

- Helmsman hears an incorrect command and performs an incorrect action.
- Helmsman hears an incorrect command and performs the correct action.

- Helmsman hears a correct command and performs an incorrect action.
- Helmsman does not acknowledge a command spoken by the Conning Officer.

This study seeks to create an experimental environment for recording each error type occurrence and calculating the ratio between the event type, subject, and trial to the total word count. The results should indicate the feasibility of using this software on a U.S. Navy warship, and specify the sources of error wherever possible.

III. METHODOLOGY AND DATA COLLECTION FOR VOICE RECOGNITION SOFTWARE EXPERIMENT

A. OVERVIEW


The objective of this study is to determine the performance of COTS speech recognition software in a simulated bridge environment. In an effort to better understand and make inferences regarding what produced, caused or contributed to SRS performance, this section presents the observational frame of reference, the assumptions and the experiment design prior to the experiment's initiation. The expectations, experiment design and possible factors reducing the reliability of the data will also be considered.

Expectations are ideas researchers have going into the experiment, which are proven true or false based on the resultant data. Each expectation considered addresses specific questions regarding software performance versus human error. The experiment is designed to reduce the chance that the outcome is due to anything but the independent variables. Note that experimental designers need to consider six major classes of information, including "post-treatment behavior or physical measurement, pre-treatment behavior or physical measurement, internal threats to validity, comparable groups, experiment errors, and the relationship to treatment". [Ref. 30]

Each of these issues will be addressed with the exception of the "comparable groups" since the experiment required individual subject comparisons, not comparisons between groups. Post-treatment relates to analysis of the data and pre-treatment considers information about all aspects of the experiment including the subjects, the software, the environment and the expectations. Internal threats to validity are factors, which discredit or make ambiguous the cause and effect relationship. Experiment errors are any actions or side effects, which result in inaccurate or false data. The relationship to treatment refers to the factors such as the sequence or setting, which may cause different effects in the data.

B. EXPERIMENT OBJECTIVES

The basic measure of performance selected in this work is the number of words not recognized divided by the total number of words on a trial run basis. Metrics include software and human errors, as described in Chapter II. Table 3, shown below, presents how the observed results are organized, where each cell lists the observation and identify the setting, simulation scenario and vessel for that trial.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Subject A	Result (c, m, d)				
Subject B					
Subject C					
Subject D					
Subject E					

C = console
S = simulator

U = underway
M = mooring
C = channel

d = Arleigh Burke (DDG)
f = Frigate (FFG)

Table 3. Experiment Expectations.

C. EXPERIMENT DESIGN

This investigation compares performance by one unit, DNSV6.0, using five subjects. The treatment was the trial performed by each subject. Each trial lasted approximately twenty to thirty minutes.

The subjects considered were in a block design group, which means that the subjects have known commonalities, which are expected to affect the outcome of the experiment. [Ref. 31] The block design applies to this study

because every subject in the group has three common properties, which are expected to affect the outcome of the experiment. The common properties are as follows: (1) extensive ship-handling experience, (2) Officer Of the Deck (OOD) qualifications and (3) male.

Factors affecting the outcome included: a) console or simulator setting, b) simulation type, underway steaming, mooring or leaving the channel and c) vessel type, destroyer or frigate. A minimum of three and maximum of five trials were performed with each subject. The trials were performed between normal Marine Safety International (MSI) operations. Therefore, some subjects executed trials one after another while other subjects completed a trial each day or when it best suited their schedule.

Randomness is important to an experiment to remove any bias, as the design of a study is biased if it systematically favors certain outcomes. [Ref. 32] Testing the subjects in varying ways decreases the likelihood that the experiment is biased. Another form of randomness introduced in our study was the difference in which simulation program and which vessel to conn was considered. The subjects had the opportunity to simulate conning an Arleigh Burke Destroyer (DDG) or a Perry-Class Frigate (FFG) with Auxiliary Power Units (APUs). Both vessels have gas turbine engines. There were three simulations to choose from a) underway steaming, b) mooring, and c) leaving the channel. There were also two locations from which to conn, at the console or in the simulator. Subjects conned from both locations. Although randomness is a positive aspect of the experiment the variation may cause experimental error.

Experimental Error is “variation produced by disturbing factors, both known and unknown”. [Ref. 33] Experimental error can lead to incorrect conclusions by data that is hidden or skewed. By reducing the unexplained variance in the experiment setting and implementation the researcher reduces the possibility of experimental error. Thus, reducing experimental error increases the probability of reaching an accurate conclusion. The design setting seeks to avoid incorrect conclusions and confusion between correlation and causation.

Correlation occurs because one or more variables are associated with another variable. For example, if there is a correlation between the type of ship, the setting and the system performance it does not mean that the system performance was directly caused by the relationship between the ship and the setting. Causation occurs when a factor produces a change in the experiment outcome. An example is the subject. The expectation is that different subjects will yield different outcomes given the same scenario or setting. Design and careful analysis will attempt to ensure each factor is appropriately seen as a cause of the result, not that the factor simply correlates with the other factors, that does not actually cause a change in the results. This leads to the complexity of effects.

Complexity of effects occurs as multiple factors are taken into consideration. The investigator must identify how the factors relate to one another, if at all, and then base a decision within those parameters. The greater the number of factors the greater chance there is for complexity of effects to occur. On a final experimental design note, this study employed the randomized block design, vice Latin square, because of potential interaction among factors.

D. EQUIPMENT AND SIMULATOR

The experiment called for the use of a laptop computer, digital recorder, and wireless microphone system. The laptop was a Fujitsu C Series LIFEBOOK with an Intel® Pentium® 4 CPU with 160 GHz and 256 MB of RAM. A Sony Digital Voice Recorder with an 8 MB Memory Stick was used to record the responses from the console operator. An operator acted as the Helmsman, Lee Helmsman and any other bridge personnel necessary for the completion of a ship's movement. A SHURE ULX/S Standard Wireless Microphone System provided the flexibility needed in a bridge environment. The ULX/S has an RF Carrier Frequency Range of 554 to 865 MHz with an effective range of 100 meters, and an Audio Frequency Response of 25 to 15,000 Hz, +/- 2 dB

variations. It uses a battery pack, which easily clips to the Conning Officer's belt or pocket. The battery life is eight to nine hours using a 9V Duracell MN1604 alkaline battery. [Ref. 34]

The experiment was performed at the MSI simulators in San Diego, California. MSI has been providing ship-handling training to the commercial maritime industry and the U.S. Navy since 1974. MSI centers utilize the latest simulation techniques to provide a realistic environment, to include the sounds associated with ship maneuvers, without real-world risks, focusing on the decision-making process vice the reaction process. Their courses are compliant with all applicable International Maritime Organization (IMO), Standards of Training, Certification and Watch-keeping for Seafarers (STCW), United States Coast Guard (USCG) and other regulations. [Ref. 35]

E. EXPERIMENT SETTING

Upon arrival at MSI the wireless system and laptop were set up at the simulator console. The console is located in an approximately 20' X 20' multi-purpose room with access to a classroom, the passageway to the simulator and the main entrance area, as shown on Figure 4. The room is used for meetings, instruction and breaks as well as the simulator command center. Foot traffic and conversations are a normal part of this setting.

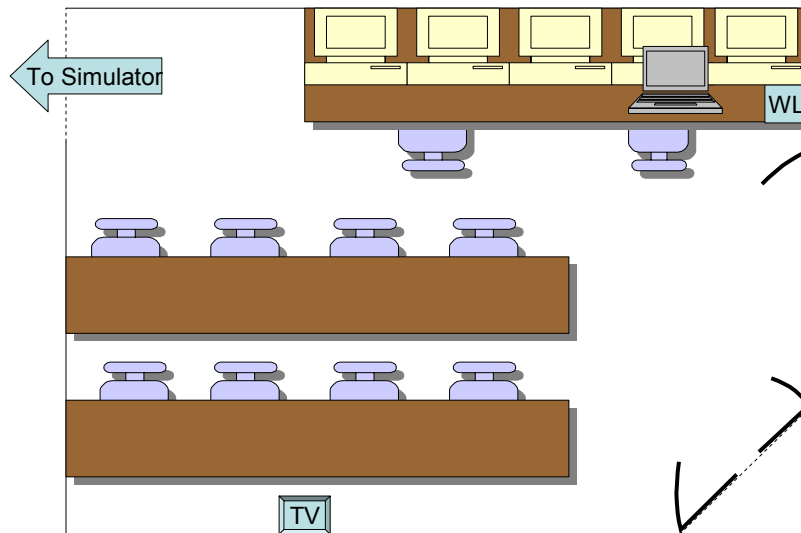


Figure 4. MSI Console Room.

The simulator is positioned approximately 50 feet away. The simulator provides a 3-D and auditory environment where Conning Officer's practice ships' movements. The simulator is significantly noisier than the multi-purpose room. Bow waves, buoy bells, environmental noise and other nautical sounds are simulated to create a more realistic environment. Table 4 below provides the noise levels in the simulator and console room throughout each type of scenario.

Bridge Readings	A weighting	C weighting	Console Readings	A weighting	C weighting
Ambient Noise (UPS and AC)	66.2 dB	70.2 dB	Ambient Noise (Computers and AC)	50.0 dB	70.0 dB
FFG Pierside (gas turbine, no wind, no bow wave)	71.1 dB	71.7 dB	FFG Pierside (gas turbine, no wind, no bow wave)	51.5 dB	71.9 dB
FFG Underway (10 Knots, 10 knot relative wind)	69.3 dB	71.7 dB	FFG Underway (10 Knots, 10 knot relative wind)	52.1 dB	73.7dB
FFG Underway (10 knots, 20 knot relative wind)	69.8 dB	72.0 dB	Doug Atherton Conning	78.1 dB	82.3 dB
FFG Underway (10 knots, 20 knot relative wind, gyro noise due to 60°/min ROT)	70.2 dB	72.6 dB	Bill Kirkland Conning	70.0 – 72.7 dB	75.9 – 76.9 dB
FFG Underway (10 Knots, 20 knot relative wind, own ship whistle)	76.7 dB	86.0 dB			
FFG Underway (10 knots, 20 knot relative wind, conning commands given)	86.3 dB	88.2 dB			
Readings were made with a sound pressure indicator.					
The voice and gyro sources were one foot from meter.					

Table 4. MSI Noise Levels.

Dragon NaturallySpeaking Version 6.0 was previously loaded into the laptop. Each participant was shown the proper positioning of the wireless

microphone headset and spent approximately 20 minutes creating a user profile, which trains the software to adjust to the speaker's speech volume, sound quality and voice. After creating the user profile a conning command vocabulary was added. Each participant trained the software to recognize the new conning commands.

F. EXPERIMENT PROCESS

After receiving an explanation of the purpose of the experiment and general guidelines for training the software, subjects fitted and adjusted the microphone to their optimal position. Next, they were asked to speak in the exact same manner as if they were giving conning commands on the ship, into the microphone, following step- by-step instructions provided in the set up of DNSV6.0 to create a user profile. Once the user profile was produced, subjects recorded a list of seamanship words and phrases into their user profile.

After creating the user profile, each subject was asked to perform a trial run in the simulator. In addition to the computer's recording, each discrepancy between the Conning Officer's speech and the software's resultant text was recorded in a narrative log. Upon completion of each trial, the data was reviewed and the original saved. A comparison of discrepancies noted in the software was followed by immediate corrections to ensure the speech engine would associate sounds with the correct words. Following the correction, a new trial was performed and the process continued. This was an iterative process where the software "learned" the user's speech patterns, and an expectation was to observe improvement with each trial run per user. [Ref. 36]

G. EXPECTATION AND CONSIDERATIONS

1. Expectations

The first assumption is that Dragon NaturallySpeaking Version 6.0 will perform differently based on the subject being studied. As discussed in Chapter

II, the subject's speech patterns, accent, software training style and voice volume affect the software accuracy. This leads to the first expectation considered in our study.

E1: Variability of software performance is dependent upon the subject.

Note that, the software is designed to learn the subject's speech characteristics after repeated use and correction, which would be indicated by an improved recognition rate. As a result, performance should see improvements with each trial, thus, leading to a second expectation.

E2: System performance will increase with subsequent trials compared to previous trials.

The setting, vessel type and simulation scenario varied among trials. Neither the vessel type nor the simulation scenario should influence the results among professional career mariners. The setting on the other hand may affect the system performance due to the difference in noise levels. These are encapsulated in the third and fourth expectations.

E3: There is no significant difference in the software performance due to the vessel type or simulation scenario.

E4: Setting affects the system performance.

Lastly, the combined effects of the subjects, simulation scenario and the setting may be a source of variation in software performance. A subject may be more comfortable conning with a particular Helmsman or in one scenario or setting, versus another. These combined interactions may influence the interpretation of the results and warrant analysis, [Ref. 37] as suggested by the fifth expectation.

E5: Interaction between the subjects, simulation scenario and setting may cause variation in the software performance.

2. Considerations

Many variables must be considered when reviewing and analyzing the results of an experiment. Each variable and its interaction with other variables affect the outcome and interpretation of the data. This section will highlight the most prominent variables.

According to the ScanSoft manufacturer, DNSV6.0 software is designed to type at least 80 percent of a user's dictation accurately after the initial training session and to achieve a 90 to 98 percent accuracy rate for most users. [Ref. 38] The expectation is that each conning officer will experience system performance at least at the stated level. The most valuable outcomes from this experiment will be regarding the software operation initially and then with repeated use.

The Helmsman and Lee Helmsman functions were performed by two individuals, each with over 30 years of ship control experience, meaning the trials probably run more smoothly than with a less experienced Helmsman. Note the human error factor regarding Helmsman performance may not necessarily be representative of the values one might observe in the fleet environment. Furthermore, the number of ship control miscues from the conning officers due to their own mistakes is anticipated to be lower because each participant has several years more conning experience than the average fleet operator. In fact, the number of errors due to misinterpretation by the Helmsman/Lee Helmsman or mistakes by the Conning Officer is expected to be rare in this environment.

The software may choose the incorrect word. There are two issues to take into account: (1) the vocabulary and (2) the statistical weighting of the vocabulary. As noted in Chapter II, DNSV6.0 has an expansive global vocabulary and allows the user to add words. Through repeated use, words were added to an individual's vocabulary, not to the global vocabulary, which is time consuming and repetitive. A ScanSoft representative pointed out a shortcoming of the SRS, which is that there is no way to add words to the global

vocabulary directly by a user. [Ref. 39] Designers must write the code explicitly defining the global vocabulary at the factory, as is done for DNS legal or medical versions.

DNSV6.0 Professional software is predefined to select the word with the highest probability of use in the typical office environment. Since seamanship terms were added to the original vocabulary for the purpose of this study, they have an extremely low statistical probability initially. Software will more likely choose a non-seamanship term until the Conning Officer uses the term enough to make it a greater statistical probability than any other word with a similar sound. For example, a Conning Officer states 'very well' in acknowledging helm responses to orders. 'Farewell' is a common closing salutation in the business world; therefore, DNSV6.0 chooses 'farewell' until 'very well' is used repetitively and corrected in the software, increasing its probability higher than that of 'farewell'.

Environment poses a challenge to the external validity of the experiment, where external validity is defined as the degree to which the conclusions reached in this study would hold for other persons, in other places and at other times. [Ref. 40] Remember, the environment for this study is not as noisy as the bridge of a ship, even though the simulator generates equipment, wind, and wave noises. In addition, there are potential internal validity issues, such as selection and experimenter bias. Internal validity is the ability to show cause and effect between dependent and independent variables. [Ref. 41] The selection factor is the extensive experience level of the participants, which tends to decrease the possibility of mistakes and misinterpretation compared with conning officers throughout the fleet. Many times the helmsman anticipates the conning officer's commands, for example. Concurrent real world operations severely limited the pool of conning officers and helmsmen available for the observation of this study. Finally, as the experiment progressed, the researchers improved ability to observe the experiment and annotate discrepancies may have lead to moderate unintentional experimenter bias.

There are many positive aspects to the study as well. The study was performed in a building with the same physical attributes as a ship, such as large metal beams and walls. These facts are comparable to a ship, realistically testing the wireless system connectivity. The wireless system allowed the participants to move about the simulator bridge as they would on a ship. Subjects exclusively used U.S. Navy standard commands in ship-handling, creating a more realistic scenario. Each candidate performed multiple trials enabling the system to learn in between trials, creating a more realistic basis for comparison. There were multiple accents and speech styles among the subjects providing a good base level of variation among participants.

H. SUBJECTS

Five subjects participated in the experiment over a five day period. None of the subjects had significant speech impediments, illnesses, or dental appliances affecting their speech. Table 5 lists the characteristics and qualifications for each subject. The asterisk denotes the Surface Warfare designation was not instituted when Subject D served in the Navy. The glossary, Appendix C, identifies the acronyms from the table.

Trait \ Subject	A	B	C	D	E
MSI Simulation Experience	Instructor 10 years	Instructor 4 years	Computer Operator 1 year	Computer Operator 9 years	No
Naval Reserves	N/A	N/A	N/A	N/A	CDR
Retired U.S. Navy	CAPT	CAPT	LCDR	OSC (E-7)	No
At Sea Command Tour	3 (1 O-5 & 2 O-6)	2 (1 O-5 & 1 O-6)	No	N/A	1 (Commercial)
Years in U. S. Navy	30	30	18	20	18
Surface Warfare Officer	Yes	Yes	Yes	*	Yes
Sea Duty	20 years	13 years	12 years	15 years	20 years
Commercial Mariner	No	No	No	No	20 years
MSI Qualifications	Ship Handling Instructor ----- ARPA, ECDIS, BRM Instructor	Ship Handling Instructor ----- ARPA, ECDIS, BRM Instructor	None	Senior Simulation Computer Operator	N/A

Table 5. Subject Traits.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA PREPARATION AND ANALYSIS OF EXPERIMENT RESULTS

A. EXPERIMENT SCENARIO

Day one, the experiment setup began by comparing the equipment onsite at the Marine Safety Institute (MSI) with the experiment equipment described in Chapter III. The MSI Technical Support Representative (TSR) noted a special connector was necessary to complete the circuit between the simulator sound system, the laptop, and the wireless microphone. Once the equipment was positioned and tested, it worked according to the manufacturer's specifications. With setup complete, the list of seamanship terms, listed in Appendix A, was added to the global vocabulary, it is the last step before the subjects began creating their profiles, as described in Chapter III.

Subject D created a new profile using the SHURE wireless microphone because he made his previous profile using a wired microphone. The need for the new profile arose when it was observed there was a difference in volume when using a wired versus wireless microphone. Subjects B, C and E created their speech profiles. The enrollment process took longer than anticipated because each subject had to record each seamanship term into individual profiles.

Day two, Subject A created a speech profile and performed the first trial. Immediately it was noticeable that the software was not recognizing the majority of words spoken, as the speaker was saturating the microphone level. Microphone volume saturation is indicated on the PC by a red line and needs to be avoided or the recorded sounds are distorted and much more difficult for the software to recognize. Subject A's first trial was stopped. The TSR verified the hardware connections were correct. After reviewing the troubleshooting chapter of the DNSV6.0 User's Guide, it was evident there was a significant difference between the subject's volume in the profile compared to the volume used in the simulator. Basically, Subject A spoke softly while reading the enrollment script

but increased his speech volume and spoke more forcefully to project his voice across the room, as if he were speaking to the helmsman in a “command voice” when giving conning commands. Note this is a common reaction for first time users and considered a form of stage fright. [Ref. 42] The user subconsciously changes speaking style because of an awareness of being recorded, but reverts to a normal speaking volume and style when in a more familiar and comfortable situation. As a result, Subject A repeated the entire enrollment process with instructions to speak in the same manner and volume as if giving commands. Subject A has a strong New York accent, which did not affect the experiment, as the results in the following trials were satisfactory and more comparable to the results of the other subjects.

Subjects B, D, and E performed a minimum of three trials each throughout the week without any noteworthy happenings. Subject C performed his first trial at the console, on the third day after several trials from Subjects A, B, and D. There were considerably fewer errors during this first trial than in any of the previous first trials. There were three possibilities for the cause of the difference, a) decreased distance between the wireless microphone and the receiver, b) noise level in the simulator versus the console room or c) Subject C spoke more clearly than the other subjects. According to the TSR, a problem with the microphone system due to the distance between the microphone and the receiver would manifest itself as dropping, not as incorrectly recognizing a word. Therefore, distance was not the problem. The answer became clearer when Subject C completed his first trial in the simulator. Subject C’s recognition rate decreased slightly in the simulator compared to the console room. The noise level in the simulator is audibly louder than at the console, decreasing the speech recognition rate. The third possibility may also have been that Subject C had a lower error rate than the other subjects regardless of scenario, setting or trial number.

B. DATA PREPARATION

The final data set consisted of 23 trials. The original data worksheets are included in Appendix B. Subjects A, B, C, and D performed five trials apiece. Subject E only performed three trials due to schedule conflicts and time constraints. Table 6 represents the raw data where the number of errors is divided by the total word count for each subject during each trial. As predicted, Subject A's first trial is drastically different from the rest of the data. This measurement may skew any statistical analysis of the data if included. The observations, described in the previous section, indicate that results for Subject A, Trial 1, might need to be removed. Aggregated error counts across software and human error types, discussed in Chapter III are the computational basis for these error rates.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Subject A	0.893	0.088	0.089	0.054	0.098
Subject B	0.061	0.110	0.080	0.083	0.052
Subject C	0.047	0.052	0.019	0.043	0.039
Subject D	0.063	0.045	0.045	0.046	0.023
Subject E	0.076	0.055	0.034		
Errors/Total Word Count					

Table 6. Raw Data Results.

1. Data Analysis Requirements

A few discussion points are necessary before heading into the data analysis. Analysis of Variance (ANOVA) is the appropriate statistical tool and requires the response variable to be normally distributed. The principle performance measure for the voice recognition system is “error” and is a zero or one response. For each word, SRS either succeeded or failed in correctly interpreting the conning commands. These are known as Bernoulli trials, which yield overall error rates as a proportion of total word count. These outcomes are distinctly non-normal because a normally distributed variable is unbounded

between negative infinity and infinity. Because the response is not normally distributed, the residuals of a basic model would also fail to meet this requirement, rendering ANOVA invalid.

Using the proportion of incorrectly interpreted words as an estimator for some unknown population parameter, θ for the probability of error in interpreting any word, the odds of failure are an adequate approach toward characterizing system performance. Equation 2 represents the odds of error.

$$\text{Odds of error} = \frac{\theta}{1-\theta}. \quad (2)$$

The *logit transform* is the inverse of the logistic function, taking its argument defined on the range $[0, 1)$ and returning output ranging from negative to positive infinity. Furthermore, taking the logarithm of the numerator and denominator yields a variable that is positive for $\theta > 0.5$ and negative for $\theta < 0.5$ and unbounded in both directions. [Ref. 43] The logit is defined as the natural logarithm of the odds of some event. The odds of an event are computed as the ratio of the probability that the event will occur divided by the probability that the event will not occur. [Ref. 44] The structure of this transformation is expressed in Equation 3 below

$$\text{logit}(\theta_i) = \log \left[\frac{\theta_i}{1-\theta_i} \right], \text{for each run } i \quad (3)$$

where the outcomes are a function of the explanatory variables based on the expectations stated in Chapter III. The logit transform yields a table of values for the log of the “odds of the SRS making an error during trial i .” These transformed values, presented in Table 7, form the basis for the data analysis and enable more appropriate use of ANOVA.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Subject A	2.1216	-2.3445	-2.3308	-2.855	-2.2208
Subject B	-2.73	-2.0868	-2.4375	-2.3979	-2.9124
Subject C	-3.0123	-2.8959	-3.9671	-3.091	-3.2155
Subject D	-2.6931	-3.0511	-3.0621	-3.0258	-3.7485
Subject E	-2.5014	-2.8478	-3.3322		

Table 7. Logit Transform Values.

2. Influential Observation

An influential observation is any case, trial in this study, whose presence causes major changes in the data results. [Ref. 45] The presence of any influential cases may become evident while investigating evidence of a normal quantile plot. A quantile plot is assumed to have a normal distribution where the data points begin in the lower left corner and follow along an imaginary straight line to the upper right corner. [Ref. 46] A plot of the overall activity as a function of subject, trial, setting and scenario yielded the following normal quantile plot shown in Figure 5. These data suggest that a singular subject's trial yielded an error rate greater than 0.5 and a positive value for the logit transform. All other points are negative, due to an error rate less than 0.5. As discussed in a previous section, the nature of this outcome was an anomaly. The resultant plot clearly demonstrates the data is not normally distributed. Furthermore, the extreme nature of this observation causes concern that it may affect the explanatory model, making it a candidate for removal as an influential observation.

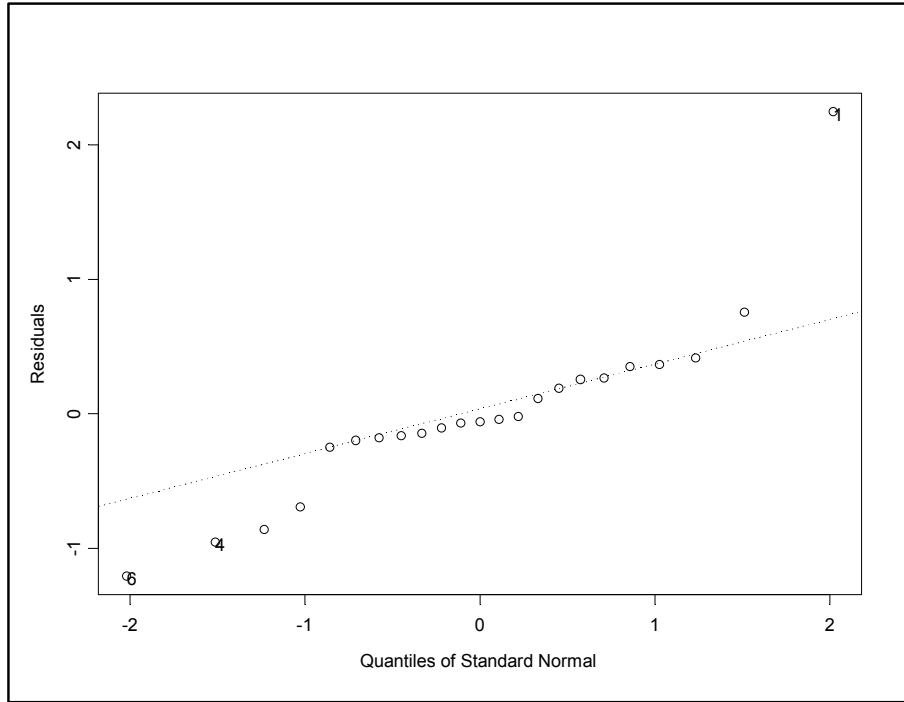


Figure 5. Quantiles of Standard Normal with Trials 1, 4, and 6 Marked as the Most Significant.

The most irregular point in this plot is the first one, Subject A, trial one. It deviates significantly from the overall pattern observations, strongly influencing the data set. This is problematic for two reasons. First, it is not characteristic of the overall performance observed throughout the rest of the experiment for the reasons already explained. Second, it will unduly alter conclusions suggested by the data set. To determine the amount of influence Subject A's first trial has on the data set the results are calculated using Cook's Distance formula. Cook's Distance is the calculation of the difference between the regression parameter with the abnormal point and the regression parameter without the abnormal point. [Ref. 47]

Essentially, Cook's Distance considers the difference in model outcomes by iteratively removing observations. Those points whose removal most markedly changes the predicted model computation yield a high value for Cook's Distance, ***D***. The greater the ***D*** value is the more substantial it changes the model, which is an undesirable situation. [Ref. 48] A graphic representation of

Cook's **D** value and its relative influence over the rest of the analysis is shown below in Figure 6. As can be seen, the problematic first observation for Subject A has by far the highest value for Cook's **D**, marked by its trial number on the plot.

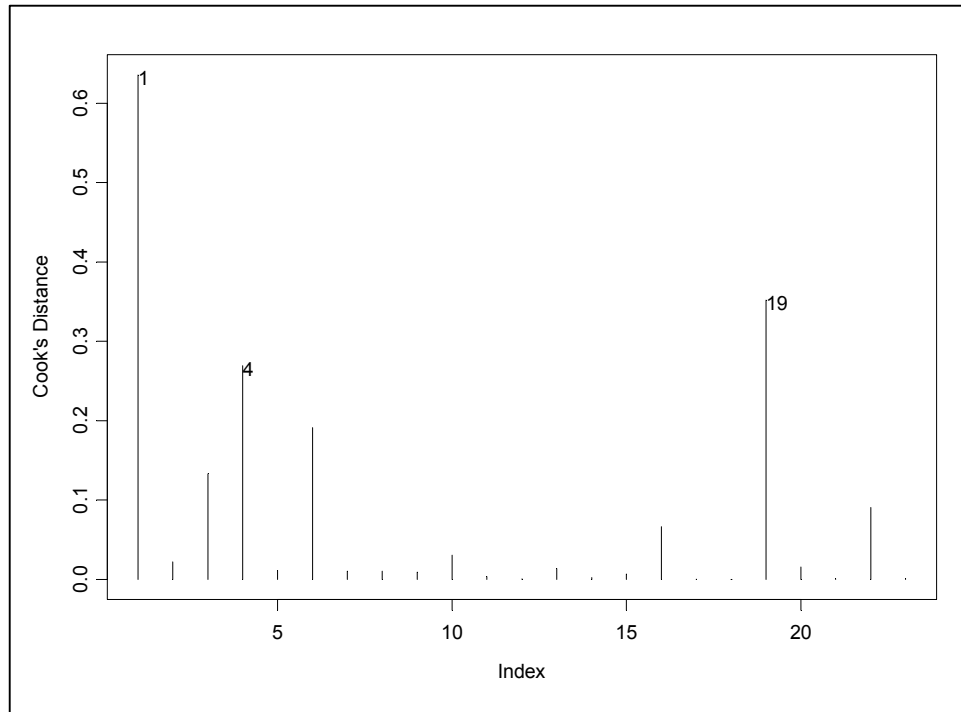


Figure 6. Cook's Distance with Trials 1, 4, and 19 Marked as Significant.

For these reasons, further analysis will omit this point, making use of a trimmed data set denoted as “tr.” in future analysis. The term ‘trimmed’ is used when labeling a table or plot to denote a data point was removed. Below in Figure 7, the Standard Normal Quantile plot shows a reasonably normal distribution for the trimmed data compared to the plot containing Subject A’s first trial. Note how the data points follow a more reasonably normal distribution without Subject A’s abnormal data point. Now that the data are more normally distributed, ANOVA may be performed on the trimmed data set.

3. Anova Methodology

The ANOVA methodology considers the role of explained and unexplained variation in performance measures as a testimony to a model's significance. Equation 4, the measure of performance, in this case SRS error is represented as follows:

$$\text{Total SRS Variation} = \text{Explained Variation} + \text{Unexplained Variation}. \text{ [Ref. 49]} \quad (4)$$

The distance between data points and their mean value measures variation. Distance is determined by squaring the mathematical difference in values. These are referred to as sums of squares, leading to the Equation 5:

$$\text{Sum of Sq (Total)} = \text{Sum of Sq (Model)} + \text{Sum of Sq (Residuals)}. \quad (5)$$

Using these sums of squares and dividing them by the appropriate degrees of freedom (Df), yields the mean square for both the model and residuals. The ratio of the mean squares is an F-statistic that measures the mean amount of variation explained by this model as compared to the mean amount of unexplained variation. To be deemed appropriate, the F-statistic requires both data sets to be normal. [Ref. 50] These data satisfy that requirement, as depicted in Figure 7.

After computing the F-statistic, based on the observed data, and comparing this value to the known F distribution, analysis yields a P-value. The P-value is the probability of observing the results seen during the experiment given that the null hypothesis is true. The null hypothesis states that introduction of an explanatory variable will not have an effect on the performance responses of the study. That is, there is no difference among model groups. This entire ANOVA methodology, including sums of squares, degrees of freedom, mean squares, F-statistic and P-values is summarized by an ANOVA table for each model associated with the five expectations identified in Chapter III.

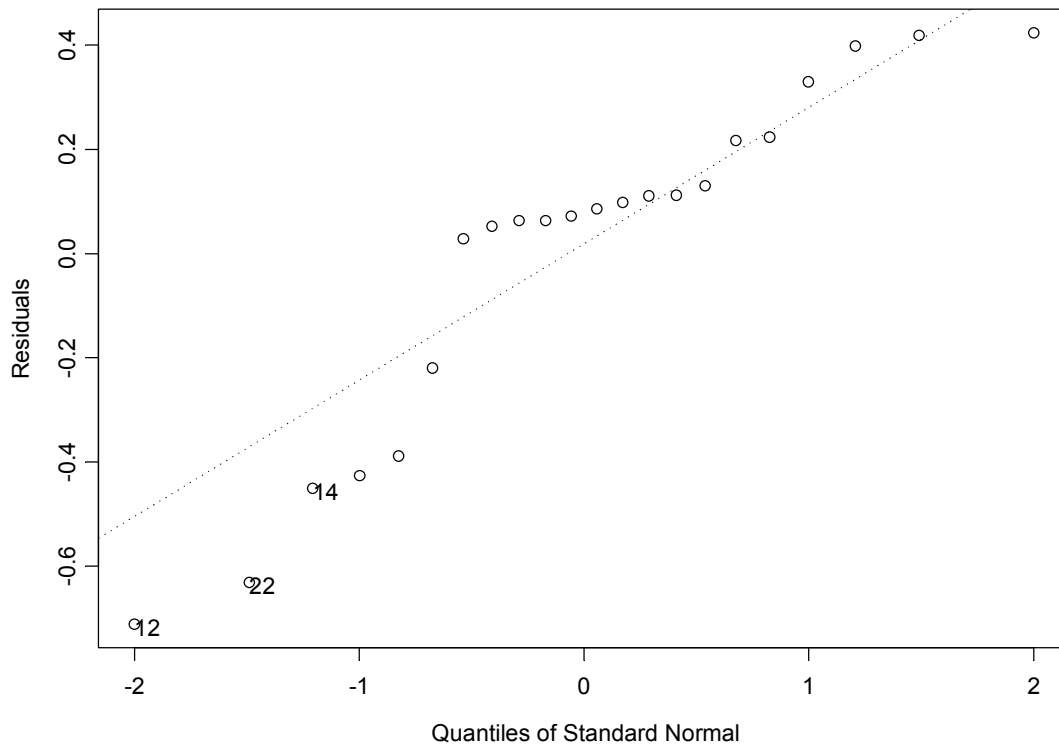


Figure 7. Trimmed Results Plot.

4. Inference Testing

a. *Expectation 1*

- **Individual Subject Accounts for Much of the Variability in Software Performance**

As noted earlier there was distinct variation among subjects' performances. A couple of the subjects' performance results were similar but other subjects performance results had several more or several fewer errors, which indicates the null hypothesis, "there is no difference in software performance due to the subject", should be rejected. The analysis of variance yielded a P-value, in Table 8, that confirms the significance of these observations.

H1	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
tr.subject	4	2.344044	0.586011	4.36397	0.01309
Residuals	17	2.282828	0.134284		

Table 8. P-Value for Expectation 1.

Another confirmation of the observations is seen in Figure 8, where each subject's performance is directly compared to another. Figure 8 shows the ninety-five percent confidence level of the difference in performance. If the data includes zero then at 95% confidence there is no distinguishable difference in performance. Note the first line A-B. These subjects overall outcomes were similar and the center point is close to zero. The 95% confidence interval includes zero, meaning there is no distinguishable difference in performance. Next when viewing A-C, the center point is skewed right to .8 and the interval does not include zero, meaning there is a distinguishable difference in performance at the 95% level of confidence.

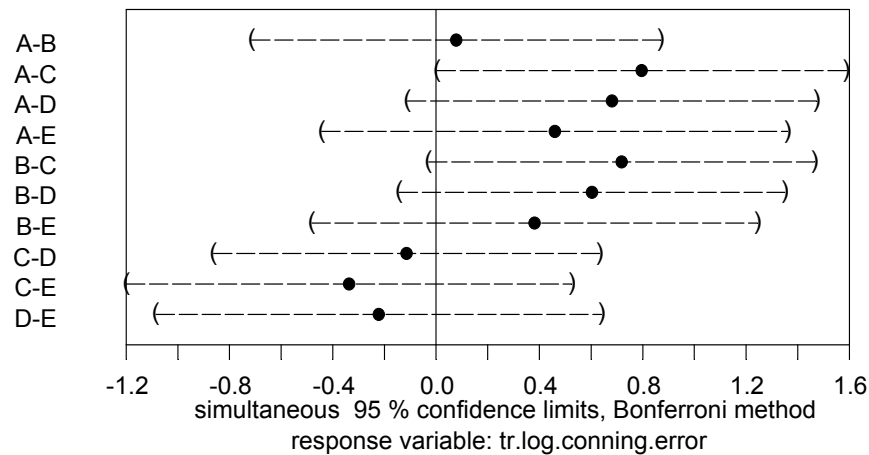


Figure 8. Subject Error Performance Similarities.

The further the comparison center is from zero, the greater the difference in the performance between the subjects. Subjects A and D performed quite differently but not as differently as Subjects A and C. Observe that Subjects A and B performed similarly so the comparison between Subject B and C is very

similar to the comparison between Subject A and Subject C. The variance between the subjects above explains over 50% of the variability in the SRS error rate. The results of this model confirm the first expectation is true.

b. Expectation 2

- **Successive Trials for Individuals Will Yield Better System Performance**

Throughout the experiment, the expectation was for the error rate to decline with each successive trial per subject. Unfortunately, those expectations were thwarted by reality. Instead, the error rate fluctuated up and down with every new trial, regardless of the subject. This was due to inconsistent enforcement of experiment controls. The subjects attempted various actions to avoid recording comments that were irrelevant to conning but important to the simulation, including turning the microphone off, trying to move it away from their mouth, and covering it. Each attempt inevitably led to a software error.

When the microphone was turned on again the subject would speak before the wireless system engaged, resulting in words not being recorded. If the subjects tried to move it or cover it up the microphone would get bumped resulting in additional words from the noise created by the contact. Other errors from contact occurred when a subject would unknowingly scratch their face, cough or rub their nose.

Another issue was the introduction of new words. The subjects introduced new vocabulary, not previously incorporated into their profiles or the global vocabulary. This led to an increased number of Software Type 2 errors “software recognizes the wrong word when the correct word is not in the vocabulary”. The P-value in Table 9 shows the strong probability that the results observed would occur given that the null hypothesis is true, thereby suggesting that trial number, above, was inconsequential.

H2	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
tr.trial	4	0.523838	0.130959	0.5426	0.70668
Residuals	17	4.103034	0.241355		

Table 9. P-Value for Expectation 2.

Figure 9 illustrates the performance comparison of the trials. The model clearly proves the trials did not improve successively, but remained relatively consistent. All the data points are clustered around zero indicating there was no distinguishable difference in performance from one trial to another. Moreover, there was no indication of positive trend looking at sequential trials. From trial 1 to 2, 2 to 3, 3 to 4, and 4 to 5, there was no consistently positive comparison of SRS response.

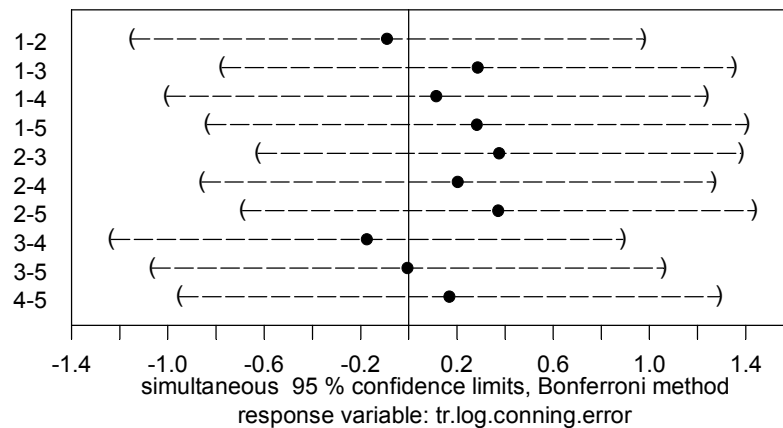


Figure 9. Trial Comparison.

c. ***Expectation 3***

- **There Is No Significant Difference in System Performance Due to Operational Scenario**

The decision to use a particular scenario or vessel in a trial varied. All three scenarios, mooring, channel, and underway, use the same commands and verbiage. The vessel type changed but it had no bearing on the study. The ambient noise between the scenarios does vary. As mentioned in Chapter III, Table 4, in the simulator, the noise level increases as the vessel moves faster.

Therefore, the noise level while leaving the channel is louder than mooring and the noise level while underway is louder than leaving the channel. Based on this information, an expectation may be to view the most errors during an underway scenario and the fewest errors during a mooring scenario. The results did not show any major differences between any of the scenarios. The P-value, shown in Table 10, indicates a 50 percent probability of observing the results observed and consequently that scenario is insignificant.

H3	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
tr.scenario	2	0.322484	0.161242	0.71174	0.50342
Residuals	19	4.304387	0.226547		

Table 10. P-Value for Expectation 3.

d. Expectation 4

- Setting Affects System Performance**

The setting, console room versus simulator, has a crucial bearing on the SRS error rate. As noted previously, the noise levels in the two locations are very different, with the simulator having considerably more ambient noise than the console room. The replicated sounds from the simulator could be heard in the console room during the underway scenario. “Dragon NaturallySpeaking® performs best in a quiet room.” [Ref. 51] The increased noise level in the simulator slightly decreased the recognition rate comparatively. After analysis, what appeared originally as a slight decrease in recognition, resulted in a substantial reduction.

	Difference in Performance	Standard Error	Lower Bound	Upper Bound
Console vs. Simulator	-0.468	0.18	-0.844	-0.0911

Table 11. Ninety-Five Percent Confidence Interval (t = 2.086).

Because this expectation is associated with comparing only two sets of data, the two-sample t-test is appropriate. [Ref. 52] Ninety-five percent

confidence interval is noteworthy in that the area encompassed by the upper and lower bounds does not include zero as exemplified in Table 11. The fact that zero is not included signifies there is a significant difference between outcomes from these two settings. The confidence interval corroborates the observations during the study, leading to the rejection of the null hypothesis; the setting does not affect the results. The upper and lower bounds equate to a difference in actual error rate between (.01, .04) as computed by an inverse of the original logit transform.

e. *Expectation 5*

- **Variation in system performance may be associated with an interaction of subject, simulation and scenario**

The last issue of concern was whether any combination of variables caused an effect of significance. The subjects were given wide latitude during testing, raising concern regarding the interaction of the variables. The subjects determined what they said, where they conned from and as remarked upon in section 3.C., the scenario and vessel used. This latitude led to further scrutiny of the data.

The original results from the first four expectations signified the need to review the possibility of interaction effects between the variables. During the study, the overall impression was that the subjects and how well they trained the system were the greatest influence on the accuracy rate. The combination of the scenario and the subject seemed like a low priority since the vocabulary was expected to remain the same for all trials. The first step considers the cumulative statistics of the Full Model, accounting for all the factors and interaction between subject and scenario. The P-value was calculated for the scenario and subject interaction. Table 11 shows the results, which was unexpectedly significant. The P-value of the subject, which is significant, is not offset by the P-value of the scenario, which is not significant, thus the null hypothesis is rejected. The combination of all the

variables plus the scenario interaction with subject, account for 80% of the variability in the SRS performance (adjusted R^2). In other words, all of these factors play a role in explaining variability in SRS performance.

	tr.trial	tr.setting	tr.scenario	tr.subject	tr.scenario:tr.subject
Sum of Squares	0.523838	0.766753	0.06032	2.211054	0.809451
Deg. of Freedom	4	1	2	4	4
	Residuals				
Sum of Squares	0.255456				
Deg. of Freedom	6				
Residual Std Error:	0.2063396				
4 out of 20 effects not estimable					
Estimated effects are unbalanced					
	Df	Sum of Sq	Mean Sq	F Value	Pr (F)
tr.trial	4	0.523838	0.1309594	3.07589	0.106258
tr.setting	1	0.766753	0.7667534	18.00903	0.0054176
tr.scenario	2	0.06032	0.0301599	0.70838	0.5294343
tr.subject	4	2.211054	0.5527634	12.98297	0.0040987
tr.scenario:tr.subject	4	0.809451	0.2023628	4.75297	0.0452828
Residuals	6	0.255456	0.042576		

Table 12. P-Value for Expectation 5 (Full Model Including Scenario-Subject Interaction).

Analyzing the interactions between the subject and setting is of great interest because they emerged as the most significant factors. The P-values from the previous single factor models indicated that both the subject and the setting are important. The question to answer is whether a subject in a particular setting provides any additional insight. If both are individually important, then perhaps the interaction between the two variables is also important. The P-value, in Table 12, of the combined variables points out that knowing which setting the subject conned from is not statistically significant, however, the addition of this variable yielded no better explanation of SRS performance. The results do not allow rejection of the null hypothesis; there is no variation in system performance due to this interaction.

	tr.subject	tr.setting	tr.subject:tr.setting	Residuals	
Sum of Squares	2.344044	0.345104	0.425363	1.512361	
Deg. of Freedom	4	1	4	12	
Residual standard error:		0.3550071			
Estimated effects may be unbalanced					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
tr.subject	4	2.344044	0.586011	4.649771	0.0169047
tr.setting	1	0.345104	0.3451036	2.738264	0.12387
tr.subject:tr.setting	4	0.425363	0.1063408	0.843773	0.5237524
Residuals	12	1.512361	0.1260301		

Table 13. P-Value for Expectation 5 (Subject * Setting).

The final model assessed the interaction between the setting and the trial. At MSI, the setting was arbitrarily chosen for any given trial. Some subjects stood in the simulator, while others stood or sat in the console room. At the time, the location was worth noting but not of interest. As evidenced by the P-value in Table 13, there is little or no significance regarding the SRS performance. This model suggests a decrease in the value of the setting as a predictor of the system execution and accounts for less than 33% of the collective variation in the SRS error rate.

	tr.subject	tr.setting	tr.trial	tr.setting:tr.trial	Residuals
Sum of Squares	2.344044	0.345104	0.421589	0.295939	1.220196
Deg. of Freedom	4	1	4	4	8
Residual Std. Error:	0.3905439				
Estimated effects may be unbalanced					
summary(tr.subject.trial.setting.aov)					
	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
tr.subject	4	2.344044	0.586011	3.842076	0.049869
tr.setting	1	0.345104	0.3451036	2.26261	0.170942
tr.trial	4	0.421589	0.1053972	0.691018	0.618499
tr.setting:tr.trial	4	0.295939	0.0739847	0.485068	0.747094
Residuals	8	1.220196	0.1525246		

Table 14. P-Value Expectation 5 (Setting * Trial).

C. EXPERIMENTAL DESIGN AND IMPLEMENTATION LESSONS LEARNED

Throughout the experiment and subsequent analysis, it became apparent improvements in the design or implementation of future experiment would yield better results. Changes in experiment implementation contributed to data unexplained variability that seemed harmless at the time, but the results made it clear the changes impacted the study. The following list contains a few of the key lessons learned.

- Begin each trial with a standard phrase to initiate the software to allow the software to engage,
- Use the speaking style appropriate to the task while creating the speech profile. This reduces errors and avoids the need to recreate the profile,
- Ensure each subject in the first trial speaks 100% of the vocabulary. Additional words unknown to the lexicon result in errors and distort the successive software learning process,
- Ensure all subjects perform the same number of trials to ensure a balanced data set for analysis,
- Wait approximately two seconds after the wireless microphone is turned on before speaking. There is a slight delay before it begins transmitting the signal to the software, resulting in error,
- Do not make contact with the microphone during recording. The software constantly seeks to create a word. Any noise activates the software and adds unwanted words to the text.
- Keep spare batteries available at all times for the microphone or invest in a rechargeable battery pack. The wireless system needs new batteries regularly. The manufacturer states the battery lasts eight to nine hours. Observe the indicator on the system to insure the battery does not die during use.
- Copy and save the original transcript prior to making corrections. The original copy contains all the errors while the corrected copy has what the conning officer actually said.

The key lessons learned about implementing an experiment are corrective actions to lessen the opportunity for disruptions or errors in future studies. Issues arose throughout the study that had not occurred during the pre-test phase, requiring small adjustments in the experiment process. For example,

when pre-testing the microphone, there had been a sufficient delay in speaking to allow the system to engage. This was not pre-planned, but occurred naturally. Also, the need for batteries may present a challenge on a ship. Rechargeable batteries are a more economical and space saving alternative. In addition there are several types of wireless microphones on the market and additional research is necessary to confirm which one is best suited for the shipboard environment.

Overall, the experiment provided useful data concerning the use of Commercial-Off-The-Shelf speech recognition software for conning ships. Improved experiment design knowledge may have resulted in a more normal data pool and led to more conclusive analysis of DNSV6.0, as numerous factors influence speech recognition software performance such as subject, trial, setting, scenario, vessel, possible Interactions, etc.

In this analysis, some interactions emerged as significant, making a randomized blocked design the most appropriate. Firm control over noise factors such as spurious verbal sounds and microphone adjustments will provide data that are more refined. However, these last two noise factors are serious characteristics of human behavior that must be considered during system design.

V. CONCLUSIONS AND RECOMMENDATIONS

A. ANALYTICAL CONCLUSIONS

This experiment was the first feasibility study for commercial-off-the-shelf (COTS) speech recognition software as a tool for conning U. S. warships and yielded important insight into SRS performance and for further studies of this system. The error rate, size of the vocabulary, and user enrollment are key design considerations in adopting this technology.

The research provides quantitative evidence that the SRS error rate is strongly dependent on the user. Users having difficulty achieving acceptable error rates are encouraged to train the software more thoroughly. The error rate is moderately impacted by the surrounding ambient noise but can be minimized by creating the user profile in the noise environment in which it is to be operated and by using noise dampening hardware.

The study emphasized the need for a focused and limited yet complete ship-handling vocabulary or lexicon. DNSV6.0 has a large vocabulary creating more opportunity for poor recognition, which is a significant drawback. It also has the ability to learn new words and to create special vocabularies, which is a positive trait. The SRS insistence on proper grammar added words and created misinterpretations in its attempt to meet the pre-defined office rules. During testing, SRS “learned” new rules required for conning within five trials.

As mentioned earlier, the user is the most significant factor in the success or failure of SRS. The user’s successful enrollment is the keystone to the process. Subject A of the study demonstrated how an erroneous enrollment can have detrimental effects on the resulting SRS accuracy rate. Users should be reminded to speak normally, using the same speech pattern, volume and speed as usually used in the specified situation.

The study also revealed some important points about the wireless microphone. Microphone position influences operational capability. The simple act of rotating the microphone upwards, toward the temple, completely stopped

speech transmission. This emphasized the high quality of the noise dampening feature built into the microphone as well as the need for correct positioning. The wireless system is power intensive, requiring frequent battery changes, but it does have an indicator letting the user know of its current status. Users attempting to use the microphone on-off switch created an unforeseen occurrence. The delay from the time the microphone was turned on until it began receiving the signal caused a lack of recognition. Once aware, the subjects did not have additional problems.

B. IMPACT OF THIS STUDY

The U. S. Navy's transformation and vision to reduce future ship size and manning requirements indicate the need for an increase in technological apparatus to perform the functions currently performed by Sailors. The Voice Activated Command System is a concept included in the design concept of the Integrated Bridge System (IBS). This concept seeks to develop technological alternatives that support safe and sound ship-handling. There are many engineering alternatives for incorporating technology and reducing manpower that preserve reliability and maintain high confidence levels but SRS is a readily available and viable option, today.

The study demonstrated that basic speech recognition software is suitable for testing and incorporation in future IBS designs. There are additional issues, which must be addressed during the design process, which were not covered in this thesis. They include the use of speaker recognition capabilities to allow certain individuals, such as the Commanding Officer; specific rights not afforded general bridge personnel. Another issue is the ability to engage and disengage the microphone. Some systems use a button while others use a keyword. The COTS SRS used in this study uses a keyword, "*microphone off*", to disengage the microphone, but the microphone must be turned on manually. This is not practical for a conning officer who must speak to bridge personnel regarding issues about the ship but not actually driving the ship. One COTS SRS

incorporates a capability for the microphone to go into a sleep or stand by mode when a key word is spoken to disengage, *“Go to sleep”* or *“Stop Listening”*. Then wait and listen for the on keyword, *“Wake Up”* or *“Listen to me”*. [Ref. 53] The words more appropriate for a ship’s bridge are “Helmsman” to activate recording and “Very well” to deactivate recording.

Speech recognition software is sufficiently technologically advanced to enable VACS to clearly receive commands from the conning officer. It is capable of recognizing and transmitting conning commands to VACS with an acceptable accuracy rate. COTS SRS is a feasible solution for achieving future Navy mission requirements.

C. RECOMMENDATIONS FOR FUTURE STUDY

The COTS SRS used for this study came straight out of the box with only one change, the addition of ship-handling vocabulary. The study did not test all features, which may have improved the results of the study. The following is a list of recommendations based on the study findings:

- Perform a follow-on study on a U. S. Navy ship to determine the potential impacts of a true ship environment and due to ambient noise differences,
- Perform follow-on trials using advanced user options. One advanced untested option was the ability to correct while speaking. In this study, all corrections were made at the end of a trial vice stopping the simulation and correcting immediately, a more effective method of improving SRS performance. Another option, which may have a profound impact, is a system which does not include a vocabulary. Current COTS SRS has such a system where a language model exists, but each individual user inserts the necessary words, such as those included in Appendix A.
- Investigate recording standard conning phrases as opposed to recording individual words during enrollment to increase recognition rates,
- Increase the time allotted to subjects during the enrollment phase to enable them to become more comfortable speaking to a computer and wearing a wireless microphone.

The results of this study indicate COTS SRS is a viable alternative for further evaluation on the high seas. As long as the components are technologically advanced and employ the best features on the commercial market, the system can support further testing. Legal and medical versions of COTS SRS prove industry has the ability to modify the system to accommodate very specific, high profile applications, and a similar approach could be followed for ship-handling operations. Specific applications require specific lexicons, meaning it only includes words necessary to complete the task. A SRS with a small, but applicable lexicon is best suited for conning operations. The smaller lexicon reduces the opportunity for the software to choose a similar yet incorrect word.

There are numerous traditional and bureaucratic reasons for not embracing a technology that does what humans have performed for centuries. However, the technology is available and ready, and the opportunity to explore change exists. Further testing and evaluation of speech recognition software to support ship control systems and processes propels ship-handling from elements employed in the days of sail and steam into the future of maneuvering warships at sea.

APPENDIX A. SHIP-HANDLING VOCABULARY

0	46	Knots
1	47	Lee Helm
2	48	Left
3	49	Magnetic
4	50	Maneuvering
5	1/3	Mark
6	2/3	Meet
7	Aft	Mind
8	Ahead	Minute
9	All	My
10	Amidships	New
11	Answers	No
12	APU	Nothing
13	APUs	Of
14	As	On
15	At	One Third
16	Automatic	Passing
17	Aye	Per
18	Back	Percent
19	Belay	Pitch
20	Bells	Port
21	Checking	Propulsion
22	Combinations	Revolutions
23	Continue	Right
24	Course	RPMs
25	Degrees	Rudder
26	Ease	Rudders
27	Emergency	Shaft
28	Engine	She
29	Engineroom	Shift
30	Engines	Sir
31	For	So
32	Full	Standard
33	Given	Starboard
34	Go	Steady
35	Goes	Steer
36	Hard	Stop
37	Head	The
38	Headings	To
39	Helm	Turns
40	Her	Two Thirds
41	How	Unit
42	Increase	Very
43	Indicate	Well
44	Is	You
45	Keep	Your

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. EXPERIMENT RESULTS

	SOFTWARE ERRORS BY TYPE				HUMAN ERROR BY TYPE				Total Word Count	S1/TWC	S2/TWC	S3/TWC	S4/TWC	H4/TWC
	1	2	3	4	1	2	3	4						
Subject A	1	2	3	4	1	2	3	4						
Trial 1	242	0	0	0	0	0	0	0	271	0.893	0.000	0.000	0.000	0.000
Trial 2	10	4	0	0	0	0	0	1	160	0.063	0.025	0.000	0.000	0.006
Trial 3	9	2	0	3	0	0	0	0	158	0.057	0.013	0.000	0.019	0.000
Trial 4	6	0	1	1	0	0	0	0	147	0.041	0.000	0.007	0.007	0.000
Trial 5	31	1	2	8	0	0	0	0	429	0.072	0.002	0.005	0.019	0.000
Subject B	1	2	3	4	1	2	3	4						
Trial 1	12	24	0	0	0	0	0	0	588	0.020	0.041	0.000	0.000	0.000
Trial 2	5	12	0	0	0	0	0	0	154	0.032	0.078	0.000	0.000	0.000
Trial 3	15	2	1	0	0	0	1	0	224	0.067	0.009	0.004	0.000	0.004
Trial 4	8	3	0	0	0	0	0	0	132	0.061	0.023	0.000	0.000	0.000
Trial 5	4	0	0	1	0	0	0	0	97	0.041	0.000	0.000	0.010	0.000
Subject C	1	2	3	4	1	2	3	4						
Trial 1	7	2	1	2	0	0	0	0	256	0.027	0.008	0.004	0.008	
Trial 2	7	1	1	1	0	0	0	0	191	0.037	0.005	0.005	0.005	
Trial 3	4	0	1	1	0	0	0	0	323	0.012	0.000	0.003	0.003	
Trial 4	6	5	1	1	0	0	0	0	322	0.019	0.016	0.003	0.003	
Trial 5	8	4	0	0	0	0	0	0	311	0.026	0.013	0.000	0.000	
Subject D	1	2	3	4	1	2	3	4						
Trial 1	22	11	0	3	0	0	0	0	568	0.039	0.019	0.000	0.005	
Trial 2	18	17	0	1	0	0	0	0	797	0.023	0.021	0.000	0.000	
Trial 3	18	13	2	2	0	0	0	0	783	0.023	0.017	0.003	0.003	
Trial 4	14	4	0	0	0	0	0	0	389	0.036	0.010	0.000	0.000	
Trial 5	20	13	2	0	0	0	0	0	1521	0.013	0.009	0.001	0.001	
Subject E	1	2	3	4	1	2	3	4						
Trial 1	18	1	1	5	0	0	0	1	330	0.055	0.003	0.003	0.015	0.003
Trial 2	10	3	0	3	0	0	0	0	292	0.034	0.010	0.000	0.010	0.000
Trial 3	5	1	0	1	0	0	0	0	203	0.025	0.005	0.000	0.005	0.000

TWC = Total Word Count	i = 1, 2, 3, 4
Si = Software Error Type	Si / TWC = Software Error Type Error Rate (%)
Hi = Human Error Type	Hi / TWC = Human Error Type Error Rate (%)

Table 15. Error Types.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Subject A	smd	smd	scd	scd	cud
Subject B	scd	smd	smd	smd	cmd
Subject C	cmd	cmd	ccd	smd	smd
Subject D	smf	suf	cud	ccd	cud
Subject E	smd	smd	cmd		
SETTING		SCENARIO		VESSEL	
S = SIMULATOR		C = CHANNEL		D = DESTROYER	
C = CONSOLE		M = MOORING		F = FRIGATE	
		U = UNDERWAY			

Table 16. Conditions Per Subject Per Trial.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. ACRONYMS

ANOVA	Analysis of Variance
AOR	Replenishment Oiler
APU	Auxiliary Power Unit
ARPA	Automatic Radar Plotting Aid
ASR	Automatic Speech Recognition
BRM	Bridge Resource Management
CAPT	U. S. Navy Rank of Captain, O-6
CDR	U. S. Navy Rank of Commander, O-5
CG	Guided Missile Cruiser
COTS	Commercial-Off-The-Shelf
CVN	Aircraft Carrier, Nuclear Propulsion
DD	Destroyer
DDG	Destroyer (Guided Missile)
DNS	Dragon NaturallySpeaking
DNSV6.0	Dragon NaturallySpeaking Version 6.0
DoD	Department of Defense
ECDIS	Electronic Chart Display and Information System

FFG	Fast Frigate (Guided Missile)
HMM	Hidden Markov Models
IBS	Integrated Bridge System
IMO	International Maritime Organization
JV 2020	Joint Vision 2020
LCDR	U. S. Navy Rank of Lieutenant Commander, O-4
LCS	Littoral Combat Ship
LPD	Amphibious Transport Dock
LST	Landing Ship, Tank
MSI	Marine Safety International
MSO	Minesweeper, Ocean
NACS	Non-Voice Activated Command System
NTR	Naval Transformation Roadmap
O-5	U. S. Navy Rank of Commander
O-6	U. S. Navy Rank of CAPTAIN
OOD	Officer of the Deck
OSC	Operations Specialist Chief

RFP	Request For Proposal
SALT	Speech Application Language Tags
SR	Speech Recognition
SRS	Speech Recognition Software
STWC	Standards of Training, Certification and Watchkeeping for Seafarers
SWO	Surface Warfare Officer
Tr.	Trimmed Data Set
TTS	Text-to-Speech
VACS	Voice Activated Command System
VAS	Voice Activated Systems
USCG	United States Coast Guard
W3C	World Wide Web Consortium
XML	Extended Mark-up Language

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- 1 United Cerebral Palsy of New York City, Environmental and Remote Control Units, [www.ucpnyc.org/info/assist/environment.cfm], as of March 7, 2003.
- 2 2003 Defense Budget Testimony: Prepared Testimony of U.S. Secretary of Defense Donald H. Rumsfeld, for the House and Senate Armed Services Committee, February 5-6 2002, [www.senate.gov/~armed_services/statement/2002/Rumsfeld.pdf], as of March 7, 2003.
- 3 NRAC, Executive Summary, Reduced Ship Manning, [nrac.onr.navy.mil], January 10, 2003.
- 4 Statement of Honorable Hansford T. Johnson, Secretary of the Navy (Acting) Before the House Armed Service Committee, February 23, 2003, [www.chinfo.navy.mil/navpalib/people/secnav/johnsonht/testimony/johnson-hasc030226.txt], as of March 7, 2003.
- 5 Flanders, R., LTJG, First Optimally Manned Ship Aces Final Evaluation Problem, Navy Newsstand, May 10, 2002, [www.news.navy.mil], as of March 7, 2003.
- 6 Department of Defense, Office of the Under Secretary for Personnel and Readiness Washington D. C., The Ninth Quadrennial Review of Military Compensation, Vol. I, p. 3, [dod.mil/prhome/qrmc/Vol1/ch1.pdf], as of March 7, 2003.
- 7 England, G., Clark, V. and Jones, J. L., Naval Transformation Roadmap: Power and Access...From the Sea, p. 5, [www.seabee.navy.mil/nmcb1/Battalion%20News/archives/Naval%20Trans%20Roadmap%20_Final_2.pdf], as of March, 1, 2003.
- 8 Shalikashvili, J., Joint Vision 2010, p. 18, [www.dtic.mil/jv2010/jvpub.htm], March 1, 2003.
- 9 Jones, D. M., Hapeshi, K. and Frankish, C., Design Guidelines for Speech Recognition Interfaces, Applied Ergonomics, 20, (1990), pp. 40-52.
- 10 Ingall's Shipbuilding, Inc., DTN Risk Reduction Test DTIB-509 Revision, September 28, 2001.

- 11 Rauterberg, M., History of HCI, Slides 107-109, [www.ipo.tue.nl/homepages/mrauterb/presentations/HCI-history.htm], as of March 7, 2003.
- 12 Terry, M., Speech in Consumer Products, Speech Technology Magazine, March/April 2003, [www.speechtechmag.com/pub/8_2/cover/1763-1.html], as of March 8, 2003.
- 13 2020 Speech Ltd., Embedded Speech Recognition Case Study: Surrey Ambulance Service, [www.2020speech.com], as of March 8, 2003.
- 14 Request For Proposal 4913-FCD0001Draft, Functional Capabilities Document for Shipboard/Classroom and Fleet Concentration Area PC Based Bridge Training Systems, April 15, 2003.
- 15 National Institute of Standards and Technology, Speech Group Home Page, [www.nist.gov/speech], as of March 9, 2003.
- 16 World Wide Web Consortium (W3C) Home Page, [www.w3c.org], as of March 9, 2003.
- 17 Laramee, F., Speech Interfaces for Games Part 1: How Speech Recognition Works, GIGnews is a Publication of GIGnews.com, Inc., Copyright 2003, GIGnews.com, Inc., [www.gignews.com/fdlspeech1.htm], as of May 19, 2003.
- 18 Zue, V., Cole, R. and Ward, W., Survey of the State of the Art in Human Language Technology, 1996, [cslu.cse.ogi.edu/HLTsurvey/ch1node4.html#parameters], as of May 19, 2003.
- 19 Zue, V., Cole, R. and Ward, W., Survey of the State of the Art in Human Language Technology, 1996, [cslu.cse.ogi.edu/HLTsurvey/ch1node4.html#parameters], as of May 19, 2003.
- 20 Ney, I. H., Human Language Technology and Pattern Recognition, [www-i6.informatik.rwth-aachen.de/web/Research/speech_recog.html], as of May 19, 2003.
- 21 Markowitz, J. A., Using Speech Recognition, p. 39, Prentice Hall PTR, New Jersey, 1996.
- 22 Wood, M., CS3421, Natural Language Engineering, Lecture 3, [www.cs.man.ac.uk/~mary/CS3421lectures/node3.html], as of May 18, 2003.

- 23 Lamel, L. F., Adda, G. and Adda-Decker, M., Designing Pronunciation Lexicons for Speech Recognition, from LIMSI 1996 Scientific Report, [m17.limsi.fr/tlp/2pg96-lex.html], as of May 19, 2003.
- 24 Mello, Jr., J. P., Dragon NaturallySpeaking 6 Professional Review, WinPlanet, February 13, 2002, [www.winplanet.com/winplanet/reviews/4059/1/], as of October 2002.
- 25 Voice Recognition Software, Consumer Search: Reviewing the Reviewers, The Wall Street Journal Online, [wsj.consumersearch.com/computers/voice_recognition_software/fullstory.html], as of September 2002.
- 26 Gao, Z. M., A Review of Dragon NaturallySpeaking 7.0, [www.imtechlaw.net/dns7_review.htm], as of May 5, 2003.
- 27 Newman, D., Dragon NaturallySpeaking 6 User's Guide, p. 9, Scansoft Inc., 2001.
- 28 Newman, D., p. 17.
- 29 Zue, V., Cole, R. and Ward, W., Survey of the State of the Art in Human Language Technology, 1996, [cslu.cse.ogi.edu/HLTsurvey/ch1node4.html#parameters], as of May 19, 2003.
- 30 [www.sytsma.com/phad530/expdesig.html], May 3, 2003.
- 31 Moore, D. and McCabe, G., Introduction to the Practice of Statistics, p. 239, Fourth Edition, W. H. Freeman and Company, New York, 2003.
- 32 Moore, D. and McCabe, G., p. 231.
- 33 Box, G., Hunter, W. and Hunter, J., Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building, p. 7, John Wiley and Sons Inc., New York, 1978.
- 34 SHURE ULX Wireless System User Guide, SHURE Incorporated, USA, 2002.
- 35 Marine Safety International, [www.marinesafety.com/sections/site/aboutus.html], as of May 5, 2003.
- 36 Box, G., Hunter, W. and Hunter, J., p. 5.
- 37 Moore, D. and McCabe, G., Introduction to the Practice of Statistics, pp. 806-811.

- 38 Newman, D., p. 161.
- 39 Derrico, T., ScanSoft Representative, Phone Conversation Dated February, 11, 2003.
- 40 Trochim, William M., The Research Methods Knowledge Base, 2nd Edition. Internet WWW Page, at URL: [trochim.human.cornell.edu/kb/index.htm], (Version Current as of June 29, 2000).
- 41 [psych.athabascau.ca/html/Validity/concept.shtml], as of May 3, 2003.
- 42 Arnold, D., Speech Recognition – Applications and Limitations for Motor Impaired, Learning Disabled, and Speech Impaired Operators, 1998 Center on Disabilities Technology and Persons with Disabilities Conference, Conference Proceedings, [www.csun.edu/cod/conf/1998/proceedings/csun98_052.htm], as of June 2, 2003.
- 43 Weisberg, S., Applied Linear Regression, 2nd Edition, (John Wiley & Sons, St. Paul, Minnesota, 1985), pp. 267-269.
- 44 Current Projects About the Sigmoid Curve, Sigmoid Curve Data Systems, [www.sigmoidcurve.com/index.html], as of June 2, 2003.
- 45 Weisberg, S., p. 118.
- 46 Levine, D., et al., Statistics for Managers Using Microsoft Excel, (New Jersey: Prentice-Hall, 1999), p. 358.
- 47 McDonald, B., A Teaching Note on Cook's Distance – A Guideline, Res.Lett.Inf.Math.Sci. (2002), Vol. 3, pp. 127-128, [www.massey.ac.nz/~wwiims/research/letters/volume3number1/macdonald.pdf], as of June 4, 2003.
- 48 Weisberg, S., pp. 118-124.
- 49 Moore, D. and McCabe, G., p. 760.
- 50 Moore, D. and McCabe, G., p. 554.
- 51 Newman, D., p. 183.
- 52 Levine, D., et al., p. 538.
- 53 Newman, D., p. 30.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Dr. Dan Boger
Chairman, Department of Information Sciences
Naval Postgraduate School
Monterey, California
4. Dr. Monique P. Fargues
ECE Dept, Code EC/Fa
Naval Postgraduate School
Monterey, California
5. LCDR Russell Gottfried
Operations Research Department
Naval Postgraduate School
Monterey, California
6. LT D. J. Tamez
Chief, Defense Infrastructure Systems
US Pacific Command, J62
Camp Smith, Hawaii
7. Mr. T. Derrico
ScanSoft, Inc.
Worldwide Headquarters
Peabody, Maryland
8. Dr. Judith Markowitz
President, J. Markowitz Consultants
Associate Editor, Speech Technology Magazine
Chicago, Illinois
9. RADM David Ramsey, USN (Ret.)
Center Director
Marine Safety International
San Diego, California
10. CAPT Stewart

Commander, Naval Surface Force, N8
San Diego, California

11. CAPT J. Kline
Chairman Warfare Innovation
Naval Postgraduate School
Monterey, California
12. Naval Information Warfare Activity
Fort Meade, Maryland
13. LCDR Raymond Buettner Jr.
Associate Professor of Information Sciences
Graduate School of Information and Operation Sciences
Naval Postgraduate School
Monterey, California
14. Paul R. Anderson
Contract Technical Representative
Pentagon Information Technology Service Center
Arlington, Virginia